
MASTERARBEIT

Herr Diplom-Ingenieur (FH)
Falko Jahn

**Untersuchung von Methoden
und Erstellung eines Pro-
grammoduls zur computer-
gestützten Berechnung von
Diffraktiven Optischen Ele-
menten (DOE)**

Mittweida, 2011

MASTERARBEIT

Untersuchung von Methoden und Erstellung eines Pro- grammoduls zur computer- gestützten Berechnung von Diffraktiven Optischen Ele- menten (DOE)

Autor:

Herr

Falko Jahn

Studiengang:

Physikalische Technik

Seminargruppe:

PT08w1-M

Erstprüfer:

Prof. Dr. rer. nat. Bernhard Steiger

Zweitprüfer:

Prof. Dr. rer. nat. Steffen Weißmantel

Einreichung:

Mittweida, 15.10.2011

Verteidigung/Bewertung:

Mittweida, 2011

MASTER THESIS

A study of methods and creation of a routine for computer aided calculation of diffractive optical elements (DOE)

author:

Mr.

Falko Jahn

course of studies:

Physical Engineering

seminar group:

PT08w1-M

first examiner:

Prof. Dr. rer. nat. Bernhard Steiger

second examiner:

Prof. Dr. rer. nat. Steffen Weißmantel

submission:

Mittweida, 15.10.2011

defence/ evaluation:

Mittweida, 2011

Bibliografische Beschreibung:

Jahn, Falko:

Untersuchung von Methoden und Erstellung eines Programmoduls zur Berechnung diffraktiver optischer Elemente (DOE). - 2011. – 10, 73, 8 S.

Mittweida, Hochschule Mittweida, Fakultät Mathematik/ Naturwissenschaften/Informatik, Masterarbeit, 2011

Referat:

Es werden die theoretischen Grundlagen zur Berechnung diffraktiver optischer Elemente von der Beugungstheorie bis zu konkreten Berechnungs-Algorithmen dargelegt. Ein in MATLAB erstelltes Programmmodul zur praktischen Berechnung wird vorgestellt und dessen Funktionsfähigkeit an einigen Beispielen demonstriert.

Inhalt

Abbildungsverzeichnis.....	VIII
----------------------------	------

Abkürzungsverzeichnis.....	X
----------------------------	---

1	Einleitung	1
1.1	<i>Motivation</i>	1
1.2	<i>Aufgabenstellung</i>	2
1.3	<i>Vorgehensweise</i>	2
1.4	<i>Danksagung.....</i>	3
2	Theoretische Grundlagen.....	5
2.1	<i>Grundlagen diffraktiver optischer Elemente.....</i>	5
2.1.1	Einführendes.....	5
2.1.2	Bauformen und physikalisches Funktionsprinzip	6
2.1.3	Anwendungen	8
2.2	<i>Beugungstheorie des Lichtes</i>	9
2.2.1	Beugung von HUYGENS bis KIRCHHOFF	9
2.2.2	FRESNEL-Beugung.....	14
2.2.3	FRAUNHOFER-Beugung.....	18
2.3	<i>FOURIER-Optik.....</i>	20
2.3.1	Grundlagen der FOURIER-Theorie.....	20
2.3.2	Verknüpfung von FOURIER-Theorie und Beugung	21
2.3.3	Ausgewählte Beispiele	24
2.3.4	Mathematische Erweiterungen.....	26
2.4	<i>Design diffraktiver optischer Elemente</i>	28
2.4.1	Grundsätzliches	28
2.4.2	Analytische Verfahren	30
2.4.3	LOHMANN-Methode.....	30
2.4.4	Technik iterativer Verschachtelung (IIT und ODIFIIT).....	36
2.4.5	Iterativer FOURIER-Transformations-Algorithmus (IFTA)	38
2.5	<i>Technische Umsetzung der Design-Verfahren.....</i>	40
2.5.1	Die MATLAB-Umgebung.....	40
2.5.2	Besonderheiten numerischer Berechnungen.....	41
2.5.3	Die Fast FOURIER Transformation (FFT).....	44

3	Umsetzung des Programmmoduls.....	49
3.1	<i>Vorbetrachtungen.....</i>	49
3.1.1	Zur Wahl des umzusetzenden Verfahrens.....	49
3.1.2	Festlegung von Eingangsparametern	49
3.1.3	Ausgabe der Ergebnisse	51
3.2	<i>Entwicklung des Programmes</i>	52
3.2.1	Grafische Benutzeroberfläche	52
3.2.2	Struktogramm.....	53
3.2.3	Anmerkungen zur Programmierung.....	55
3.2.4	Verbesserungspotential.....	56
4	Theoretische Verifizierung der Ergebnisse	59
4.1	<i>Testergebnisse.....</i>	59
4.1.1	Strahlformung: Top-Hat-Profil.....	59
4.1.2	Strahlformung: Linien in x- und y-Richtung	61
4.1.3	Strahlformung: Hochschullogo.....	62
4.1.4	Strahlformung: Buchstabe M	63
4.1.5	Strahlteiler: 3x3-Array.....	63
4.2	<i>Überprüfung mit VirtualLab 4.....</i>	65
4.3	<i>Auswertung</i>	68
4.3.1	Anwendung Strahlformung	68
4.3.2	Einfluss des Eingangsstrahlradius	69
4.3.3	Anwendung Strahlteilung.....	71
5	Zusammenfassung.....	73
Literatur		75
Anlagen		79
Anlage 1 – Struktogramm des FFT-Algorithmus.....		A-1
Anlage 2 – Code des MATLAB-Programmmoduls		A-3
Anlage 3 – Bilder		A-8
Selbständigkeitserklärung.....		A-9

Abbildungsverzeichnis

Abbildung 1 Beugungsmuster eines Amplitudengitters, bestrahlt mit gaußverteiletem Laserstrahl.....	6
Abbildung 2 diverse DOE-Bauformen	7
Abbildung 3 DOE-Anwendungsbeispiele Strahlformung und -teilung	8
Abbildung 4 Darstellung des HUYGENSSchen Prinzips.....	10
Abbildung 5 Grundlage des KIRCHHOFFSchen Beugungsintegrals.....	11
Abbildung 6 Winkelabhängigkeit des Neigungsfaktors	13
Abbildung 7 Gültigkeit der verschiedenen Beugungsbeschreibungen	15
Abbildung 8 Koordinatensystem für FRESNEL-Beugung.....	16
Abbildung 9 Bildentstehung bei FRAUNHOFER-Beugung	19
Abbildung 10 Rechteckfunktion und ihre FOURIER-Transformierte.....	24
Abbildung 11 Variation der Spaltbreite der Rechteckfunktion	25
Abbildung 12 DIRACsche Deltafunktion und ihre Eigenschaften	26
Abbildung 13 Varianten von DOE-Anordnungen	29
Abbildung 14 Prinzip der DOE-Diskretisierung in der LOHMANN-Methode	31
Abbildung 15 Ergebnisse eines mit der Lohmann-Methode berechneten DOEs.....	35
Abbildung 16 Verschachtelung von zwei Unter-DOEs.....	36
Abbildung 17 Prinzip des IFT-Algorithmus	39
Abbildung 18 Periodizität von Funktionen bei der DFT	46
Abbildung 19 GUI des entwickelten MATLAB-Moduls	52

Abbildung 20 Struktogramm der Funktion "Set_Matrix"	54
Abbildung 21 Struktogramm der Funktion IFTA.....	55
Abbildung 22 Gaußverteilung als simulierter Eingangsstrahl	59
Abbildung 23 Simulierte Ausgangsverteilung konvergiert zum gewünschten..... Strahlprofil	60
Abbildung 24 Simuliertes Ergebnis: Top-Hat-Profil	60
Abbildung 25 Simuliertes Ergebnis: Linieneprofil in x- und y-Richtung	61
Abbildung 26 Simuliertes Ergebnis: Hochschullogo.....	62
Abbildung 27 Simuliertes Ergebnis: Strahlformung "M"	63
Abbildung 28 Simuliertes Ergebnis: 3x3-Strahlteiler-Array.....	64
Abbildung 29 Strahlprofil des Fan-Out-Elements.....	64
Abbildung 30 Screenshot: Lightpath-Diagramm des verwendeten Modells in..... VirtualLab 4	65
Abbildung 31 mit VirtualLab 4 simulierte Intensitätsverteilungen	66
Abbildung 32 mit VirtualLab 4 simulierte Intensitätsverteilung des Hochschullogos.....	67
Abbildung 33 simuliertes Ergebnis: Halbkugelverteilung	68
Abbildung 34 simuliertes Ergebnis: Strahlformung "M" bei verschiedenen Eingangsradien	69
Abbildung 35 mit VirtualLab simuliert: Linienprofil bei verschiedenen..... Eingangsstrahlradien.....	70
Abbildung 36 simulierte Ergebnisse: verschiedene Strahlteiler-Arrays	71

Abkürzungsverzeichnis

in alphabetischer Reihenfolge

CGH	Computer-Generiertes Hologramm (engl. Computer generated hologram)
DFT	Diskrete FOURIER Transformation
DOE	Diffraktives Optisches Element (engl. diffractive optical element)
FFT	Fast FOURIER Transformation („schnelle FOURIER-Transformation“)
GUI	Graphical User Interface („grafische Benutzeroberfläche“)
IFTA	Iterativer FOURIER-Transformations-Algorithmus (engl. iterative FOURIER-Transform Algorithm)
IIT	Iterative Interlacing Technique („Technik iterativer Verschachtelung“)
MATLAB	Matrix Laboratory
ODIFIIT	Optimal Decimation-in-frequency Iterative Interlacing Technique („Technik iterativer Verschachtelung mit optimierter Frequenzabschwächung“)
POCS	Projection onto convex sets („Projektion auf konvexe Mengen“)

1 Einleitung

1.1 Motivation

Diffraktive optisch Elemente (DOE) gehören zu den innovativsten Entwicklungen optischer Bauelemente der letzten Jahre. Binäre diffraktive Optiken ermöglichen die Fokussierung von Wellen in bislang schwer zugänglichen Wellenlängenbereichen, diffraktive Linsen gestatten Fokusedurchmesser unterhalb der wellenoptischen Auflösungsgrenze und spezielle Phasentransmissions-Elemente erlauben hoch effektive anwendungsorientierte Strahlformungen. Diese Beispiele und viele weitere Anwendungen aus Gebieten wie optische Abbildungen, Strahlteilung oder –homogenisierung werden durch diffraktive Optiken verbessert oder teilweise erst möglich.

Obwohl diesem speziellen Thema großes Potential innewohnt, ist die verfügbare Literatur überschaubar, die insbesondere die theoretischen Grundlagen beleuchtet. Von besonderem Interesse sind die Methoden und Vorgehensweisen, um die teilweise sehr komplexen Strukturen zu ermitteln, die für die oben genannten Anwendungen nötig sind. Häufig werden jedoch nur einzelne spezielle Verfahren behandelt und erklärt. Eine vollständige Übersicht über mögliche Lösungsansätze wäre wünschenswert.

Die meisten Optikdesign-Programme, die in Industrie und Forschung genutzt werden, sind inzwischen in der Lage, auch den Strahlengang durch ein vorgegebenes DOE zu simulieren. Eine Recherche nach entsprechender Software, die darüber hinaus die Berechnung der notwendigen DOE-Struktur ermöglicht, führte dagegen zu nur wenigen Treffern. Ein Beispiel ist das Programm VirtualLab 5 der Firma LightTrans, welches eine Vorreiterrolle in diesem Bereich einnimmt.

Die Lizenzkosten für dieses Programm liegen allerdings im Bereich mehrerer tausend Euro pro Jahr. Zudem lassen sich die Entwickler dieser kommerziell genutzten Software nicht in die Karten sehen, so dass eine konkrete Nachverfolgung des Lösungsweges schwer möglich ist. Dies jedoch ist für wissenschaftliches Arbeiten unerlässlich.

Aus diesen Gründen ergibt sich die Notwendigkeit, einerseits eine geordnete, nachvollziehbare und geschlossene Theorie der DOE-Berechnung zusammenzustellen. Andererseits sollte eine eigene DOE-Berechnungssoftware entwickelt werden, deren Rechenwege jederzeit nachvollziehbar sind.

1.2 Aufgabenstellung

Ziel dieser Arbeit ist die Entwicklung eines Programmmoduls, das die Berechnung diffraktiver optischer Elemente ermöglicht.

Dies beinhaltet zunächst die ausführliche Aufarbeitung der zugrunde liegenden Theorie als wichtige Grundlage der Arbeit. Besonderer Wert wird auf die exakte mathematische Ausführung gelegt.

Die verschiedenen möglichen Berechnungsverfahren sind auf ihre Eignung für die MATLAB-Umgebung zu überprüfen und gegebenenfalls zu bewerten. Ein geeigneter Algorithmus ist zur Umsetzung im Programm auszuwählen.

Das Programmmodul soll unter MATLAB erstellt werden und eine benutzerfreundliche, grafische Bedienoberfläche besitzen. Es soll die Eingabe von Eingangsparametern ermöglichen. Dabei ist zuvor nach praktischen Gesichtspunkten festzulegen, welche Eingangsparameter in welcher Form abgefragt werden sollen.

Im Modul sind Möglichkeiten zu implementieren, verschiedene gewünschte Ausgangsintensitätsverteilungen festzulegen. Denkbar ist beispielsweise eine Importfunktion von Bilddateien oder eine Auswahlfunktion aus bestimmten Standard-Verteilungen.

Die ausgearbeiteten theoretischen Algorithmen sind in MATLAB zu implementieren, so dass aus den Eingangswerten die Berechnung der DOE-Strukturen erfolgt. Die Ausgabe der Ergebnisse ist in geeigneter Form bzgl. Darstellung und Dateiformat zu wählen.

Zur Verifizierung der DOE-Berechnungen dient die aktuell vorhandene Optikanalyse-Software, mit der die Ergebnisse nachgerechnet und bestätigt werden. Ist eine praktische Herstellung einer ermittelten DOE-Struktur im Rahmen der Masterarbeit möglich, kann auch dies zur Verifizierung der Berechnungsmethoden herangezogen werden.

1.3 Vorgehensweise

Die vorliegende Masterarbeit dokumentiert die schrittweise Erfüllung der genannten Aufgaben. Die einzelnen Etappen in der Realisierung werden in den nachfolgenden Kapiteln beschrieben.

Im ersten Schritt sind die theoretischen Grundlagen der Berechnung ausführlich auszuarbeiten. Von den Grundlagen der Lichtbeugung und FOURIER-Optik über verschiedene Verfahren zum DOE-Design bis hin zu den Besonderheiten der numerischen Berechnung werden die elementaren Zusammenhänge ausführlich in Kapitel 2 „Theoretische Grundlagen“ beschrieben.

Auf Basis dieses theoretischen Fundaments sind die Algorithmen auszuwählen, die für eine MATLAB-Umsetzung am geeignetsten sind. Kapitel 3 befasst sich mit der anschlie-

ßenden Entwicklung des Programms. Es wird die Erstellung der Struktogramme erläutert, die wichtige Ausgangspunkte für die Programmierung darstellen. Der Quelltext des entwickelten Programmmoduls ist im Anhang einsehbar.

Im Anschluss an die Erstellung des MATLAB-Moduls werden die Beispiele der simulierten Endverteilungen vorgestellt. Hierfür werden einige in der Lasertechnik typische Strahlprofile, wie z.B. ein Top-Hat-Profil oder ein Linienprofil, erstellt. Anschließend ist eine Verifizierung der erhaltenen Berechnungsergebnisse notwendig. In diesem Schritt, mit dem sich Kapitel 4 befasst, werden die ermittelten DOE-Strukturen mit rechentechnischen Methoden kontrolliert. Dies erfolgt über den Vergleich mit anderen Programmen, wie z.B. LightTrans VirtualLab 4. Die Berechnung der resultierenden Feldverteilungen mittels bereits an der Hochschule erstellter Programme sowie die Berechnung von Strukturen, deren Ergebnisse bereits bekannt sind, könnten zusätzliche Bestätigung bringen.

Die erhaltenen Ergebnisse werden in Kapitel 5 zusammengefasst und in einem kurzen Ausblick werden Anregungen für Weiterentwicklungen gegeben.

Jeder Abschnitt der praktischen Arbeit beginnt mit wenigen kurzen, separaten Sätzen, die den Inhalt und die Erkenntnisse des Abschnittes zusammenfassen. Im Anhang ist der vollständige Quelltext des programmierten MATLAB-Moduls zu finden sowie alle Ergebnistabellen von Berechnungen und Messungen.

1.4 Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die mir die Erstellung dieser Arbeit möglich gemacht haben und die mir bei der Niederschrift auf verschiedenste Weise behilflich gewesen sind.

Zunächst sind da meine beiden betreuenden Professoren, Herr Professor Dr. rer. nat. Bernhard Steiger und Herr Professor Dr. rer. nat. Steffen Weißmantel, zu nennen. Ihre fachlich wertvolle Unterstützung und die konstruktiven Kritik motivierten mich stets auf Neue, den gestellten Anforderungen gerecht zu werden. Besonders für die gewährten Freiheiten und das entgegengebrachte Vertrauen möchte ich mich bedanken.

Auch allen Kollegen, die mir mit Rat und Tat zu Seite gestanden haben, möchte ich hiermit meinen Dank. Vor allem Markus Olbrich möchte ich für die vielen fachlichen Ratschläge und Ideen danken. Ich hoffe, mich dann im Gegenzug bei seiner Masterarbeit revanchieren zu können.

Natürlich sind auch alle Freunde, Verwandte und Bekannte nicht zu vergessen, die mich im entsprechenden Zeitraum ertragen haben und mir trotzdem stets unterstützend Mut gemacht haben. In erster Linie möchte ich meiner lieben Anne danken, dass sie die ganze Zeit an meiner Seite war und mich aufgebaut hat, wenn ich am verzweifeln war. Auch all

denjenigen möchte ich danken, die sich die Mühe gemacht haben, sich durch den Wust meiner Arbeit zu kämpfen auf der Jagd nach Orthografie- und Grammatikfehlern.

Meinen Eltern danke ich für die aufmunternden Worte, meiner Schwester für die Massagen, wenn mir mein Rücken mal wieder die viele Computerarbeit übelnahm. Und auch bei Freunden, wie Silke und Daniel möchte ich mich bedanken, dass sie durch ihr beharrliches Nachfragen nach meinen Fortschritten stets das mahnende schlechte Gewissen wach gehalten haben.

Falko Jahn,

Mittweida, 2011

2 Theoretische Grundlagen

2.1 Grundlagen diffraktiver optischer Elemente

2.1.1 Einführendes

Der Begriff „diffraktives optisches Element“ umfasst ganz allgemein alle optischen Bauteile, deren Wirkungsweise auf dem physikalischen Prinzip der Beugung (Diffraktion) von Licht beruht. Im einfachsten Fall ist schon der Doppelspalt dazuzuzählen, den Thomas YOUNG 1801 für sein bekanntes Experiment zum Beweis des Wellencharakters von Licht verwendete.¹

Mit Ausnahme von einigen Anwendungen, die die dispersiven Eigenschaften beugender Elemente (z.B. Beugungsgitter in Spektrometern) ausnutzen, galt die Beugung lange Zeit als negative Randerscheinung. Besonders in der abbildenden Optik begrenzen Beugungseffekte die maximal erreichbare Auflösung². Der deutsche Physiker und Optiker Ernst ABBE gilt als federführend bei der Untersuchung dieser Auflösungsbegrenzung.

Erst mit Aufkommen der Holografie Mitte des 20. Jahrhunderts und der Entwicklung von computergenerierten Hologrammen und Kinoformen in den 60er und 70er Jahren wurde damit begonnen, den Effekt der Beugung bewusst für optische Komponenten einzusetzen.³ In diesem Zusammenhang fallen häufig die Begriffe „binäre Optiken“ oder „Holografische optische Elemente“.

Diese Begriffe werden in den verschiedenen Literaturstellen zum Teil unterschiedlich verwendet, da die Grenzen zwischen den jeweiligen Elementen fließend ineinander übergehen. Die Einteilung in dieser Arbeit orientiert sich an den Ausführungen von [O'Shea], demzufolge „Diffraktives optisches Element“ der Überbegriff für die genannten Bauelemente ist. „Holografische optische Elemente“, optisch durch Interferenz erzeugte Beugungsmuster, und kinoforme Elemente mit kontinuierlicher Phasentransmissions-Funktion⁴ sollen für die folgenden Betrachtungen keine Rolle spielen. Auch werden Beugungsgitter in ihren verschiedenen Verbreitungsformen

¹ [Lipson], S. 3

² vgl. [Pedrotti], S. 485-490

³ vgl. [Ersoy], S. 1

⁴ vgl. [O'Shea], S. 2

(Amplituden-, Phasen- und geblazete Gitter) eine deutlich untergeordnete Rolle spielen.

Das Hauptaugenmerk der nächsten theoretischen und praktischen Ausführungen liegt auf der Berechnung der verschiedenen Formen von unregelmäßigen DOEs. Diese werden vorwiegend als „binäre Optiken“ ausgeführt, da die Herstellungsmethoden, die an der Hochschule Mittweida möglich sind, zu den typischen diskreten Höhenstufen führen. Die gewünschten DOEs entsprechen am ehesten „Computer-generierten Hologrammen (CGH)“. Nach [O'Shea] besteht der größte Unterschied zu „einfachen“ DOEs darin, dass bei CGHs auch noch höhere Beugungsordnungen erscheinen können.⁵

2.1.2 Bauformen und physikalisches Funktionsprinzip

Entsprechend der Vielzahl unterschiedlicher Bezeichnung existieren viele Erscheinungsformen dieser Elemente. Gemeinsam haben alle nur die optische Beugung als Basis ihres Funktionsprinzips.

In Abbildung 1 ist das typische Beugungsmuster eines regelmäßigen Amplitudengitters dargestellt. Die einzelnen Beugungsordnungen sind deutlich sichtbar. Ihre Lage und Verteilung ist bekanntermaßen von den Parametern des Gitters abhängig.

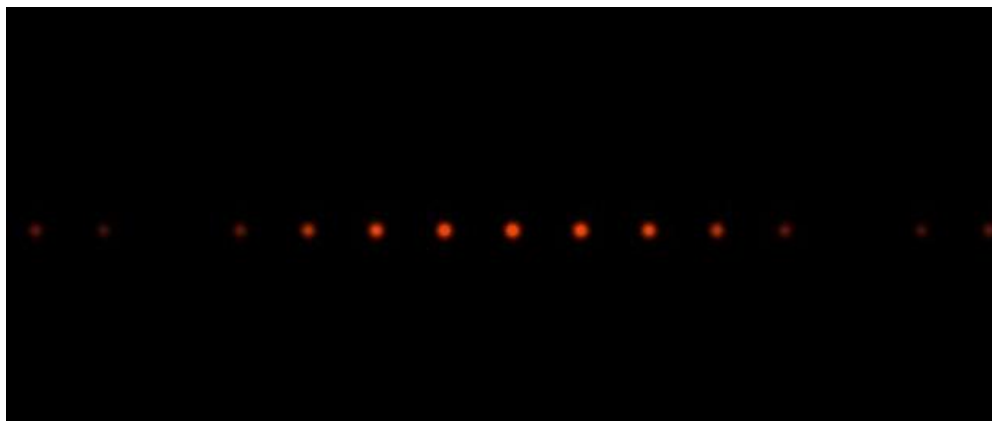


Abbildung 1 Beugungsmuster eines Amplitudengitters, bestrahlt mit gaußverteiletem Laserstrahl
[simuliert mit Software VirtualLab 4]

Wird die Regelmäßigkeit des Gitters unterbrochen, indem z.B. einzelne Gitterspalte willkürlich verschlossen werden, hat dies Einfluss auf die Beugungsordnungen. Ei-

⁵ vgl. [O'Shea], S. 2

nige Ordnungen werden aufgrund von destruktiver Interferenz geschwächt werden, während andere Ordnungen verstärkt werden.

Der Grundgedanke diffraktiver optischer Elemente ist es, diese Interferenzeffekte so gezielt durch Manipulation des „Gitters“ zu steuern, dass ein beabsichtigtes Bild entsteht. In den meisten Fällen wird diese Manipulation das regelmäßige Gitter in eine eher unregelmäßige Anordnung verändern. Dieser Gedanke ist keineswegs auf Amplitudengitter beschränkt. Auch die gezielte Manipulation eines Phasengitters hat eine solche Wirkung. Im Extremfall könnten sogar beide manipuliert werden, was jedoch in der Praxis schwierig herzustellen ist.

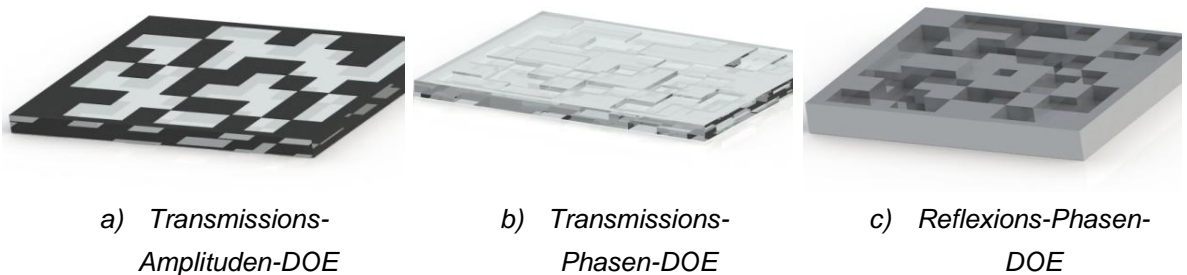


Abbildung 2 diverse DOE-Bauformen [Illustration erstellt mit Solid Works]

Abbildung 2 zeigt verschiedene Möglichkeiten, wie ein solches diffraktives optisches Element praktisch ausgeführt sein kann. Neben den schon genannten Varianten, Amplitude oder Phase in Transmission zu manipulieren, ist ein fiktives Reflexions-DOE schematisch dargestellt. Auch diese Bauform kann prinzipiell als Phasen- oder Amplituden-DOE realisiert werden. Im ersten Fall sind die zu erzeugenden Höhenunterschiede des Profils i.d.R. deutlich größer als für die Transmissions-Varianten, da die Phasenverschiebung im umgebenden Medium (meist Luft mit $n = 1$) erzeugt werden muss. Im zweiten Fall wird auf der reflektierenden Oberfläche des DOE eine absorbierende Schicht erzeugt.

Dies stellt nur einen Ausschnitt aus den denkbaren Bauformen dar. Diffraktive Elemente können z.B. auf der Oberfläche herkömmlicher refraktiver Elemente deren optische Eigenschaften beeinflussen. Auch sind die oben abgebildeten Profile und Amplitudenmasken mit kontinuierlichen Übergängen denkbar. Dies ist allerdings auch fertigungstechnisch schwerer zu realisieren. Die verschiedenen Herstellungsverfahren von DOEs werden in dieser Arbeit nicht thematisiert.

Zusammengefasst kann die Wirkung eines DOEs mit der Modulation einer einfallenden Welle beschrieben werden. Diese Modulation verändert – je nach Bauform – die Amplitude, die Phase oder beides so gezielt, dass die resultierende Wellenfront in einer Zielfläche ein genau gewünschtes spezifisches Beugungsbild ergibt. Es muss nun geklärt werden, wie diese Modulation ermittelt werden kann.

In den folgenden Abschnitten soll die komplexe Theorie, die den DOEs zugrunde liegt, in überschaubaren Etappen aufgeschlüsselt und dargelegt werden. Das quali-

tativ erklärte Funktionsprinzip soll dabei auch auf ein mathematisch stabiles Fundament gestellt werden. Die grundlegenden Begriffe und Prinzipien der Optik, wie z.B. Beugung, Interferenz und Superposition, werden allerdings als bekannt vorausgesetzt.

2.1.3 Anwendungen

Es existiert ein breites Spektrum von Anwendungsmöglichkeiten für DOEs, das hier lediglich kurz umrissen werden kann. Aufgrund ihrer Flexibilität bzgl. des erzeugbaren Beugungsmusters werden sie besonders häufig zur Strahlformung und -teilung eingesetzt.

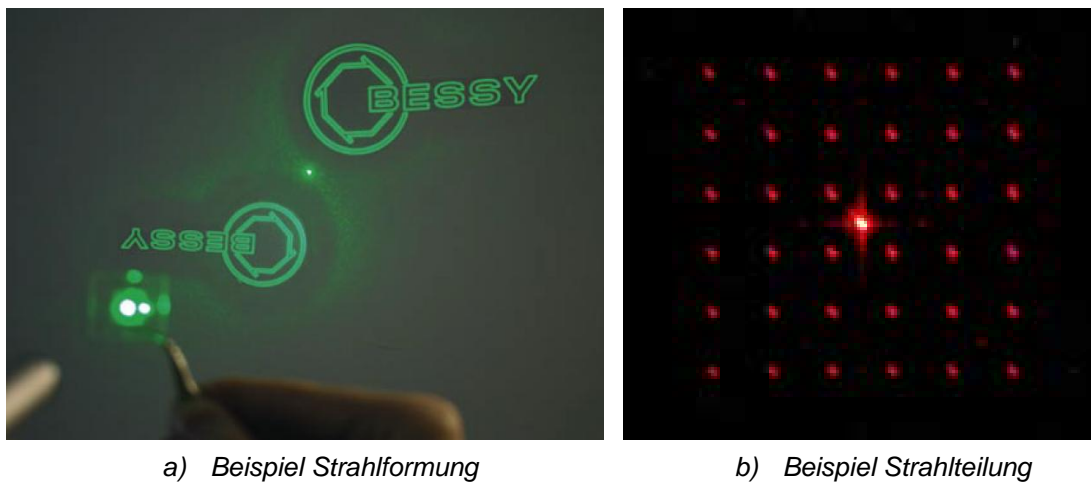


Abbildung 3 DOE-Anwendungsbeispiele Strahlformung und -teilung
(Quellen: [Löchel] und [Stankovic], S. 51)

Wie Abbildung 3 zeigt, können realisierte Strahlformer im Bereich Multimedia und Werbung verwendet werden. Zielmarkierungen und Fadenkreuz-Abbildungen lassen sich für Justage nutzen. Werden die DOEs auch für hohe Leistungen ausgelegt, können individuelle Leistungsverteilungen z.B. für die Lasermaterialbearbeitung generiert werden. Auch Strahlhomogenisierer z.B. für Excimerlaser sind möglich⁶.

Strahlteilungs-DOEs, sogenannte Fan-Out DOEs, ermöglichen eine gleichzeitige Beleuchtung mehrerer Punkte. Dies kann in Messanwendungen zum Einsatz kommen oder auch mit genügend hohen Leistungen beim parallelen Laserbearbeiten von Oberflächen an verschiedenen Stellen⁷. Abbildung 3 zeigt rechts das Beu-

⁶ vgl. [Völkel]

⁷ vgl. [Kaakkunen]

gungsbild eines beleuchteten 6x6 Fan-Out DOEs. Es ist auch möglich, das DOE so zu berechnen, dass bestimmte Punkte ausgelöscht und andere verstärkt werden.

Da diffraktive Elemente deutlich leichter und kompakter ausgeführt werden können als komplexe Aufbauten aus konventionellen Optiken, besteht auch Potential bei Miniaturisierung von Optiken (z.B. CD/DVD-Geräte).⁸ In Verbindung mit herkömmlichen Optiken zu sogenannten Hybrid-Optiken ermöglichen sie eine Korrektur von chromatischen und Abbildungsfehlern. Auch hier können aufwändige Korrektur-Optiksysteme aus mehreren refraktiven Elementen gespart werden⁹.

Darüber hinaus werden DOEs mittlerweile in vielen anderen Bereichen eingesetzt. Sie ermöglichen höhere Datenraten bei optischen Datenträgern durch gleichzeitige Verwendung mehrerer Wellenlängen¹⁰. Einige verbessern die direkte Lichteinkopplung in planare Wellenleiter¹¹. Andere ermöglichen als FRESNELSche Zonenplatte die Fokussierung elektromagnetischer Wellen, für die es sonst keine refraktiven Materialien gibt (z.B. Röntgenstrahlen)¹². Speziell mikrostrukturierte Oberflächen können sogar als Anti-Reflex-Schichten für optische Bauelemente wirken.

2.2 Beugungstheorie des Lichtes

2.2.1 Beugung von HUYGENS bis KIRCHHOFF

Mitte des 17. Jahrhunderts beobachtete der italienische Physiker Francesco Maria GRIMALDI die wichtigste Eigenschaft von Licht, die sich nicht mit der bis dahin allgemein anerkannten Theorie der Teilcheneigenschaft erklären ließ – die Beugung an einem Spalt.¹³ Der englische Universalgelehrte Robert HOOKE bestätigte einige Jahre später diese Entdeckung und führte erstmals die Idee der Welleneigenschaft von Licht ein, die der Niederländer Christian HUYGENS in seinem 1690 veröffentlichten Werk „Tractatus de lumine“ (dt.: Abhandlung über das Licht) zur Erklärung verschie-

⁸ vgl. [Brinkmann]

⁹ vgl. [P'Shea], S. 3

¹⁰ vgl. [Pawlowski]

¹¹ vgl. [Fricke]

¹² vgl. [Pedrotti], S. 539 - 542

¹³ vgl. [Lipson], S. 3-5

dener optischer Phänomene ausbaute.¹⁴ Sein entwickeltes HUYGENSSches Prinzip der Wellenausbreitung wird auch heute noch zur Erklärung der meisten Welleneigenschaften von Licht herangezogen.

Dieses Prinzip besagt sinngemäß, dass jeder Punkt einer Lichtwelle Ausgangspunkt einer kugelförmigen Elementarwelle ist. Die Einhüllende der einzelnen Elementarwellen ergibt dann die neue Wellenfront. Dabei beschränkte HUYGENS die Elementarwellen auf die Vorwärtsbewegung, so dass die geradlinige Ausbreitung von Licht und die Ausbildung scharfer Schatten damit erklärt werden konnte:

„Um nun zu den Eigenschaften des Lichtes überzugehen, bemerke ich zuerst, dass jeder Wellentheil sich in der Weise ausbreiten muss, dass seine äussersten Grenzen immer zwischen den nämlichen vom leuchtenden Punkt aus gezogenen geraden Linien bleiben. Der Wellenteil BG, [...] welcher den leuchtenden Punkt A zum Mittelpunkt hat, wird sich also bis zu dem von den Geraden ABC, AGE begrenzten Bogen CE ausbreiten.“¹⁵

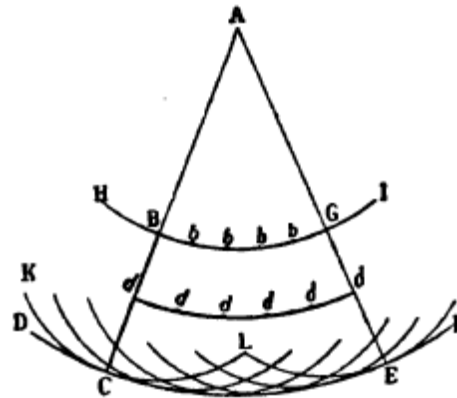


Abbildung 4 Darstellung des HUYGENSSchen Prinzips (aus [Huygens], S.24)

Abbildung 4 aus HUYGENS' Veröffentlichung deutet auch die Erklärung für die Beugungserscheinungen an Kanten an. Da er jedoch offenbar die Ergebnisse von GRIMALDI nicht kannte, ließ er selbst diesen Punkt unberücksichtigt.¹⁶

Etwa 1818 vervollständigte der Franzose Augustin Jean FRESNEL HUYGENS' Theorie um das Prinzip der Superposition. Demnach können sich Elementarwellen nicht nur verstärken, sondern auch auslöschen. Er löste damit die ungeklärte Frage, warum die Elementarwellen nur vorwärtsgerichtet eine Einhüllende bilden und führte erstmals eine Richtungsfunktion ein, die die Elementarwellen nach ihrer Abstrahlrichtung wichtet¹⁷.

Circa 70 Jahre später gelang dem deutschen Physiker Gustav KIRCHHOFF eine exaktere Formulierung. Obgleich er in seiner Herleitung streng mathematisch von einem Randwertproblem der elektromagnetischen Wellengleichung ausging, führte

¹⁴ vgl. [Pedrotti], S. 20

¹⁵ [Huygens], S. 24

¹⁶ [Born], S. 143

¹⁷ vgl. [Born], S. 143-147

diese auf eine Integralform, die HUYGENS' Prinzip der kugelförmigen Elementarwellen beinhaltet. Dieses Integral, häufig als HUYGENS-KIRCHHOFF-Beugungsintegral oder in manchen Quellen¹⁸ auch als FRESNEL-KIRCHHOFF-Beugungsintegral bekannt, ist für die Beugungstheorie grundlegend.

Ausgangspunkt ist eine punktförmige Lichtquelle Q (siehe Abbildung 5), deren elektromagnetische Strahlung über eine ebene Blende R zum Beobachtungspunkt P gelangt. Die an der Stelle S auf die Blende R treffende Welle kann als Kugelwelle beschrieben werden:

$$E_S \sim \frac{E_q}{r_s} e^{ik_0 r_s} \quad (2-1)$$

$E_S \dots$	elektr. Feldstärke im Punkt S	$E_q \dots$	elektr. Feldstärke der
$r_s \dots$	Abstand von Q zu S		Quelle

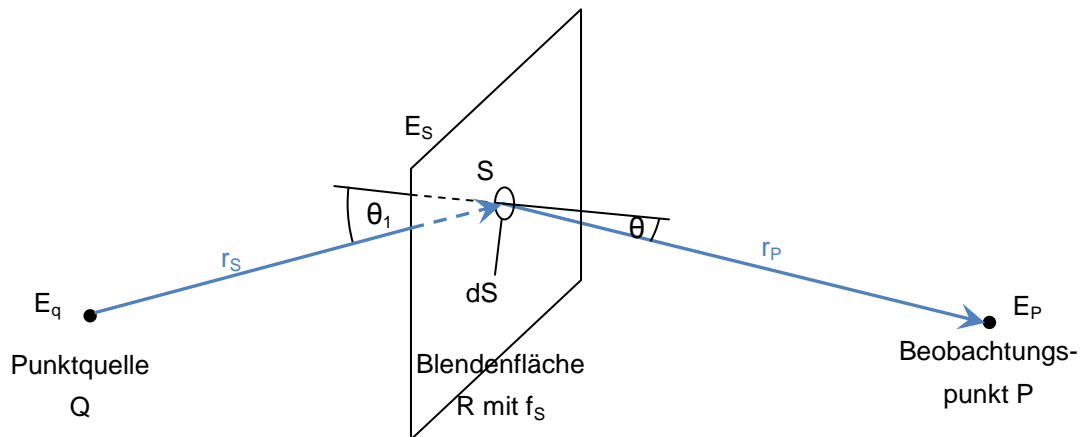


Abbildung 5 Grundlage des KIRCHHOFFSchen Beugungsintegrals

Wie in allen folgenden Betrachtungen, wird auch hier die Herleitung auf die elektrische Feldstärke E beschränkt werden. Mittels der bekannten MAXWELLSchen Gleichungen lässt sich an jedem Punkt aus dem elektrischen auch das magnetische Feld berechnen.

Gemäß dem HUYGENSSchen Prinzips wird am Punkt S durch E_s ein infinitesimal kleines Flächenelement dS zum Aussenden einer ebenfalls kugelförmigen Elementarwelle angeregt. Die emittierte Elementarwelle muss zur einfallenden Welle eine konstante Phasenbeziehung besitzen, damit das Beugungsmuster keine zeitliche Abhängigkeit erfährt¹⁹. Die Sekundärwelle strahlt mit der Gesamtstärke:

¹⁸ siehe [Pedrotti], S. 529 ff

¹⁹ vgl. [Lipson], S. 151

$$E_s' \sim E_s \cdot f_s dS \quad (2-2)$$

und fügt damit am Punkt P den Anteil dE zur Gesamtfeldstärke E bei:

$$dE \sim \frac{E_s'}{r_p} e^{ik_0 r_p} = \frac{E_s \cdot f_s}{r_p} e^{ik_0 r_p} dS \quad (2-3)$$

Der Phasenterm $\frac{1}{r_p} \cdot e^{ik_0 r_p}$ in (2-3) deutet den Kugelwellencharakter der Sekundärwelle an. f_s ist die Transmissionsfunktion der Blende R an der Stelle S. Sie kann im einfachsten Fall die Werte 0 in den Stegen und 1 in den Lücken für perfekte Transmissionsgitter annehmen. Darüber hinaus kann f_s auch komplexe Werte zur Darstellung beliebiger Transmissionsgrade und Phasenverschiebungen annehmen. Schwerpunkt im weiteren Verlauf dieser Arbeit werden Verfahren sein, die eine Berechnung dieser Transmissionsfunktion ermöglichen. Im Punkt P überlagern sich nach HUYGENS' Prinzip alle Teilwellen dE zur Gesamtfeldstärke E , was durch Integration über die Blendenfläche R dargestellt wird:

$$E \sim \iint_R \frac{E_s \cdot f_s}{r_p} e^{ik_0 r_p} dS$$

$$E \sim E_q \iint_R \frac{f_s}{r_s \cdot r_p} e^{ik_0 (r_p + r_s)} dS \quad (2-4)$$

Die exakt mathematische Herleitung von KIRCHHOFF zeigt, dass das Integral noch mit einem konstanten Proportionalitätsfaktor $\frac{-i \cdot n \cdot k_0}{2\pi}$ multipliziert werden muss, um ein mathematisch konsistentes Ergebnis zu erhalten. Dieser Faktor war ebenfalls von FRESNEL in seinen Überlegungen hergeleitet worden. KIRCHHOFF bestätigte diese dann mathematisch und konnte auch die von FRESNEL gemachten Annahmen auf Grundlage der Wellengleichung begründen. Im vollständigen **HUYGENS-KIRCHHOFF-Beugungsintegral**²⁰

$$E = -\frac{i \cdot n \cdot k_0 \cdot E_q}{2\pi} \iint_R \frac{f_s}{r_s \cdot r_p} e^{ik_0 (r_p + r_s)} \cdot N dS \quad (2-5)$$

ist zudem noch die Brechzahl n des Mediums und der Neigungsfaktor N enthalten. Dieser ist ähnlich aufzufassen, wie die Richtungsfunktion K von FRESNEL. Er berücksichtigt das winkelabhängige Abstrahlverhalten der Elementarwellen und ist wie folgt zu berechnen:

$$N(\theta, \theta_1) = \frac{\cos \theta + \cos \theta_1}{2} \quad (2-6)$$

²⁰ vgl. [Pedrotti], S. 530

Die Winkel θ_1 und θ liegen jeweils zwischen der Oberflächennormale von R und dem einfallenden Quell- bzw. dem emittierten Streustrahl (siehe Abbildung 5). (2-6) ist zu entnehmen, dass N von beiden Winkeln gleichermaßen abhängt. Einleuchtender wäre eine Abhängigkeit von der Differenz der beiden Winkel. Es stellt sich jedoch heraus, dass die Emission auch von der Lage der Oberfläche R abhängig ist.²¹

Abbildung 6 zeigt die Winkelabhängigkeit des Neigungsfaktors in Polarkoordinaten. Die einfallende Welle trifft dabei unter $\theta_1 = 0^\circ$ auf die Fläche R. In Abbildung 6 ist zudem der Neigungsfaktor, wie er von HUYGENS impliziert wurde, abgetragen. Vergleichend ist festzustellen, dass im Paraxialgebiet bei sehr kleinem θ und für die Rückwärtsbeugung bei $\theta = 180^\circ$ beide Modelle übereinstimmen.

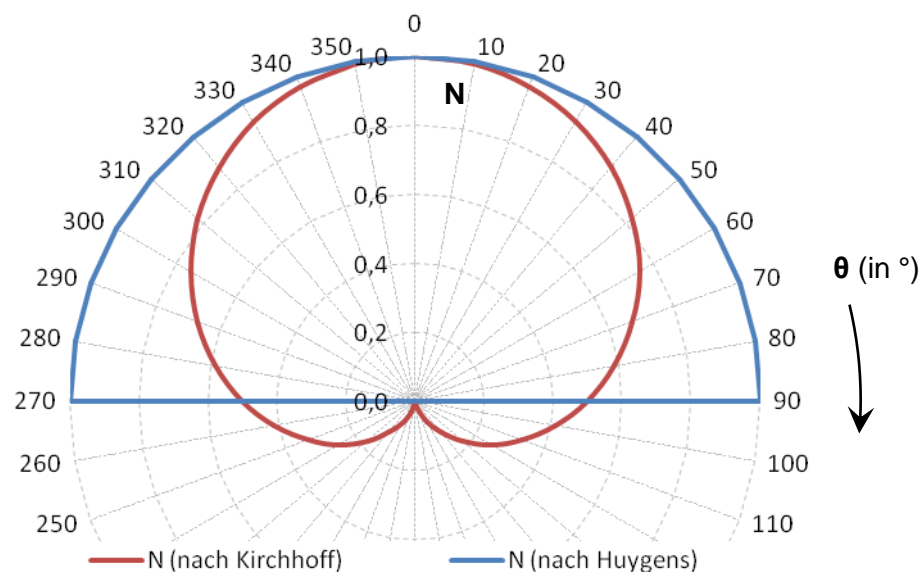


Abbildung 6 Winkelabhängigkeit des Neigungsfaktors

Es ist zu beachten, dass die elektrische Feldstärke E generell eine vektorielle Größe aus x -, y - und z -Komponente ist. Eine physikalisch ganz exakte Betrachtung muss die einzelnen Komponenten und damit den Polarisationszustand der elektromagnetischen Wellen berücksichtigen. Die bisher getroffene Vereinbarung eines skalaren E -Feldes darf jedoch unter der Annahme beibehalten werden, dass die Wellenausbreitung in einem homogenen, isotropen Medium stattfindet. In einem solchen Medium findet keine Wechselwirkung zwischen x - und y -Komponente statt, so dass die oben aufgestellten Gleichungen für beide unabhängig voneinander gelten²². Aus diesen beiden Komponenten kann die z -Komponente gemäß der MAXWELLSchen Gleichungen berechnet werden.

²¹ vgl. [Lipson], S. 155

²² vgl. [Wyrow2006], S. 55

Der Vollständigkeit halber soll hier erwähnt werden, dass durchaus Randbedingungen denkbar sind, bei denen der Polarisationszustand der Wellen Einfluss auf die Beugungserscheinungen hat. Wird Licht beispielsweise am Spalt eines perfekt leitenden Metallbleches gebeugt, wird direkt auf der Spaltoberfläche die zur Oberfläche senkrecht polarisierte Feldstärkekomponente bevorzugt. Da sich dieser Effekt jedoch auf Bereiche in der Größenordnung einer Wellenlänge bewegt, muss er auch nur für so kleine Spaltdimensionen berücksichtigt werden.²³

2.2.2 FRESNEL-Beugung

Chronologisch betrachtet, müssen die Erkenntnisse, die der französische Physiker Augustine Jean FRESNEL zu Beginn des 19. Jahrhunderts über die Natur der Beugung gewonnen hatte, noch vor KIRCHHOFFs Beugungsintegral eingeordnet werden. Tatsächlich hatte FRESNEL bereits eine abgewandelte Form jenes Integrals durch eine intuitive Herleitung erhalten, die ihm die Berechnung von Beugungsbildern ermöglichte. Dabei hatte er verschiedene Näherungen und Annahmen getroffen, die KIRCHHOFF später auch mathematisch exakt herleiten und begründen konnte.

Die Recherche nach der FRESNEL-Beugung in verschiedenen Fachbüchern zeigt mehrere Herleitungs- und Erklärungsansätze auf, die sich z.T. auch deutlich voneinander unterscheiden. Ein Weg ist das Nachvollziehen der Überlegungen von FRESNEL, verbunden mit einer eher intuitiven Herleitung der Gültigkeitsgrenzen²⁴. Die andere Variante basiert auf der Tatsache, dass die FRESNEL-Beugung ein Sonderfall der KIRCHHOFF-Beugung ist und somit auch im bereits beschriebenen Beugungsintegral enthalten ist. Diese Herleitungen führen bestimmte Vereinfachungen ein, mit denen sich aus dem KIRCHHOFFschen Beugungsintegral das FRESNELsche Beugungsintegral ergibt. In diesem Zusammenhang wird auch häufig von der FRESNEL-Näherung gesprochen.

Das bedeutet, dass generell alle Beugungserscheinungen, abgesehen vom absoluten Nahfeld wenige Wellenlängen hinter der Beugungsapertur, mittels des KIRCHHOFF-Integrals berechnet werden können. Es existiert aber ein Bereich, ab dem jene Vereinfachungen vertretbar sind und eine vereinfachte Rechnung möglich ist – der Bereich der FRESNEL-Beugung.

In Abbildung 7 wird dieser Zusammenhang verdeutlicht. Darüber hinaus illustriert die Abbildung zwei weitere Besonderheiten. Zum einen existiert mit zunehmender Entfernung von der Apertur A ein zusätzlicher Bereich, in dem weitere Vereinfachungen

²³ vgl. [Lipson], S. 150

²⁴ vgl. [Pedrotti], S. 533-534

chungen zulässig sind. Diese FRAUNHOFER-Näherung wird in 2.2.3 ausführlicher betrachtet werden.

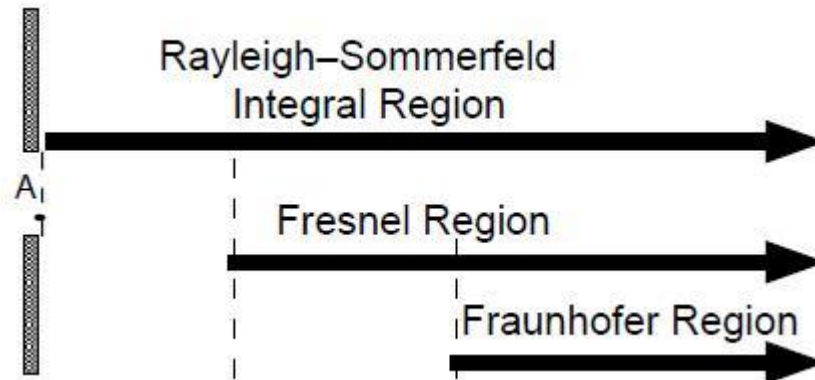


Abbildung 7 Gültigkeit der verschiedenen Beugungsbeschreibungen
(Quelle: [Ersoy], S. 64)

Zum anderen wird statt des KIRCHHOFF-Beugungsintegrals das Integral von RAYLEIGH und SOMMERFELD angegeben. Arnold SOMMERFELD, ein deutscher Physiker und Wegbereiter moderner mathematischer Methoden in der Physik, war mit einigen Voraussetzungen, die KIRCHHOFF und FRESNEL angenommen hatten, nicht einverstanden²⁵. So ist unter anderem die Festlegung, dass die Feldgrößen auf der Oberfläche der Apertur Null sind, nur unter der Annahme akzeptabel, dass der Abstand des Beobachtungspunktes von der Apertur viel größer als die Wellenlänge ist. Er änderte KIRCHHOFFs Ansatz leicht durch Anwendung einer anderen GREENSchen Funktion ab und leitete das allgemeine RAYLEIGH-SOMMERFELD Beugungsintegral her, welches auch in unmittelbarer Nähe zur Blende gültig ist.

Mit der erwähnten Vereinfachung $\lambda \ll |\vec{r} - \vec{r}_0|$, die für Lichtwellen aufgrund der extrem kleinen Wellenlänge fast immer erfüllt sein wird, ergibt sich die RAYLEIGH-SOMMERFELDSche Beugungsformel:

$$E(\vec{r}) = -\frac{i \cdot n}{\lambda} \iint_{-\infty}^{\infty} E_0(\vec{r}_0) \cdot \frac{e^{ik|\vec{r}-\vec{r}_0|}}{|\vec{r} - \vec{r}_0|} \cdot \frac{z}{|\vec{r} - \vec{r}_0|} dx_0 dy_0 \quad (2-7)$$

Abgesehen vom geänderten Integrationsbereich, dem anders formulierten „Neigungsfaktor“ und der allgemeineren Feldverteilung $E_0(\vec{r}_0)$ in der Beugungsebene anstelle von Kugelwellen entspricht (2-7) der Lösung KIRCHHOFFs²⁶ (2-5). Unter Berücksichtigung der genannten Bedingungen dürfen beide Gleichungen verwendet werden.

²⁵ vgl. [Sommerfeld]

²⁶ vgl. [Träger], S. 126

Auf Grundlage des bereits hergeleiteten KIRCHHOFF-Beugungsintegrals (2-5) wird im Folgenden die FRESNEL-Näherung hergeleitet. Stellenweise wird die Variante des alternativen, intuitiven Herleitungsweges angedeutet. Das Beugungsintegral wird für die FRESNEL-Beugung mit allgemeiner Feldverteilung $E_0(\vec{r}_0)$ in der Beugungsebene (keine Kugelwellen) geschrieben, da häufig von planaren einfallenden Wellen ausgegangen wird²⁷. Es soll das in Abbildung 8 dargestellte Koordinatensystem gelten:

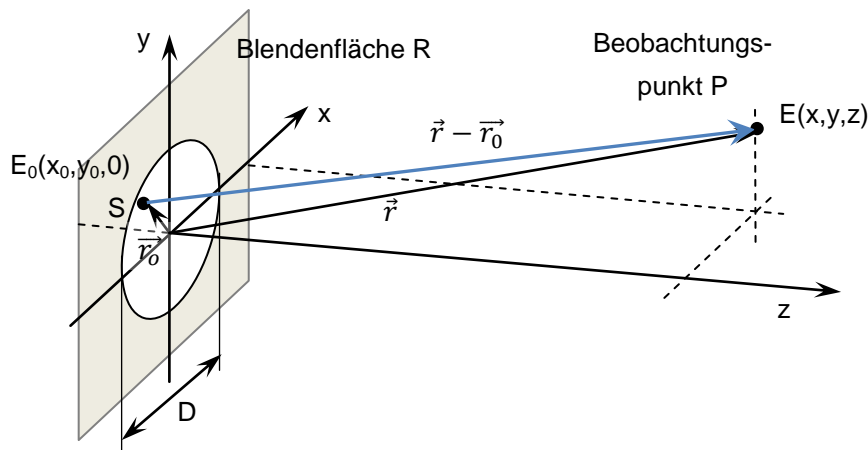


Abbildung 8 Koordinatensystem für FRESNEL-Beugung

Ab einer bestimmten Entfernung z der Bildebene von der Aperturbene R sind folgende Vereinfachungen möglich:

- $D \ll z$
- $|\vec{r} - \vec{r}_0| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - 0)^2}$

Über verschiedene Umformungen und die Entwicklung in eine TAYLOR-Reihe, deren höhere Glieder vernachlässigt werden dürfen, ergibt sich:

$$|\vec{r} - \vec{r}_0| \cong |\vec{r}| + \frac{x_0^2 + y_0^2}{2|\vec{r}|} - \frac{x \cdot x_0 + y \cdot y_0}{|\vec{r}|}$$

- Steht $|\vec{r} - \vec{r}_0|$ im Nenner, genügt der 1. Term der Näherung und es gilt:

$$|\vec{r} - \vec{r}_0| \cong |\vec{r}|$$

- Aufgrund des Paraxialgebietes ist der Neigungsfaktor $N \approx 1$.
- $|\vec{r}|$ kann im Paraxialgebiet noch weiter vereinfacht werden:

$$|\vec{r}| = z \cdot \sqrt{1 + \frac{x^2 + y^2}{z^2}} \approx z + \frac{x^2 + y^2}{2z}$$

Außer im Exponenten der e-Funktion darf der zweite Summand vernachlässigt werden.

²⁷ vgl. [Lipson], S. 156

sigt werden, so dass gilt:

$$|\vec{r}| \approx z \quad \text{und} \quad \frac{1}{|\vec{r}|} \approx \frac{1}{z}$$

Werden all diese Vereinfachungen in (2-5) umgesetzt, ergibt sich nach einigen Umformungen und der Berücksichtigung konstanter Terme das FRESNELsche Beugungsintegral:

$$E(\vec{r}) = -i \frac{n}{\lambda \cdot z} \cdot e^{i \frac{2\pi n \cdot z}{\lambda}} e^{i \pi n \frac{x^2 + y^2}{\lambda \cdot z}} \iint_A E_0(\vec{r}_0) \cdot e^{i \pi n \frac{x_0^2 + y_0^2}{\lambda \cdot z}} \cdot e^{-2\pi i n \frac{x \cdot x_0 + y \cdot y_0}{\lambda \cdot z}} dx_0 dy_0 \quad (2-8)$$

Diese Formulierung ist mathematisch günstiger als die von KIRCHHOFF, wie später noch gezeigt werden wird. Von besonderem Interesse ist nun natürlich noch der minimale Abstand, ab dem sie genutzt werden darf. Das heißt, wo ist die FRESNEL-Näherung gültig?

Vor allem diese Kriterien variieren stark in der Literatur. Bei [Pedrotti] wird zur Charakterisierung eine FRESNEL-Zahl eingeführt. Eine grobe Regel besagt dabei, dass für Werte zwischen 10^{-2} und 10^2 FRESNEL-Beugung vorliegt.

[Lipson] entscheidet die Frage nach der Beugungsart physikalisch einleuchtender mit der Änderung der Phase $k \cdot |\vec{r} - \vec{r}_0|$ bei Bewegung des Punktes S in der Apertur A. Ändert sich die Phase der Welle linear, ist FRAUNHOFER-Beugung erlaubt, während nichtlineare Änderungen das FRESNEL-Beugungsintegral erfordern. Auch diese Anschauung führt zur FRESNEL-Zahl, gegeben durch

$$N_F = \frac{D^2}{4 \cdot \lambda \cdot z} \geq 1 \quad (2-9)$$

Diese Definition ist vergleichbar mit der hier angewandten Einteilung nach [Träger]. Entscheidend für das zu verwendende Integral ist, ob die zuvor getroffenen Vereinbarungen und Vereinfachungen möglich sind. Die vernachlässigten höheren Terme der TAYLOR-Reihe für $|\vec{r} - \vec{r}_0|$ bestimmen im Wesentlichen die Grenze. Der erzeugte Gangunterschied durch diese Terme darf $\lambda/2$ nicht überschreiten.

Diese Forderung mündet letztendlich in einem Fehlerfaktor, der nach Möglichkeit sehr klein gegen 1 sein soll:

$$Q_{Fresnel} = \frac{n \cdot D^4}{16 \cdot \lambda \cdot z^3} \ll 1 \quad (2-10)$$

2.2.3 FRAUNHOFER-Beugung

Auch Joseph von FRAUNHOFER, ein deutscher Optiker und Physiker, beschäftigte sich mit den Hintergründen der Beugung. Die von ihm 1821 erzeugten Beugungsmuster enthielten eine weitere Vereinfachung der FRESNELSchen Theorie.²⁸ Diese Effekte werden heute als FRAUNHOFER- oder Fernfeld-Beugung bezeichnet.

Charakteristisch für diesen Bereich der Beugung sind die ebenen Wellenfronten der einfallenden und gebeugten Wellen. Theoretisch setzt dies voraus, dass sich Quelle Q und Beobachtungspunkt P in unendlicher Entfernung befinden. Praktisch wird diese Bedingung durch den Einsatz von Sammellinsen realisiert. Wird Q in den Brennpunkt einer solchen Linse platziert, können durch Kollimation nahezu ebene Wellenfronten erzeugt werden. Analog dazu bildet eine Linse im Bildbereich das Beugungsbild statt im Unendlichen in der Brennebene ab.

Ebene Wellenfronten bedeuten auch, dass sich die Struktur des erhaltenen Beugungsbildes mit Änderung der Position z nicht mehr ändert. Lediglich die Größe variiert in Abhängigkeit von z²⁹. Es bietet sich somit eine Beschreibung über Winkel an.

In Übereinstimmung mit [Träger] werden Richtungskosinus in x- und y-Richtung eingeführt:

$$\alpha = \frac{x}{|\vec{r}|} \qquad \beta = \frac{y}{|\vec{r}|} \qquad (2-11)$$

Der Abstand von Beobachtungs- und Beugungspunkt $|\vec{r} - \vec{r}_0|$ kann somit geschrieben werden als:

$$|\vec{r} - \vec{r}_0| \cong |\vec{r}| + \frac{x_0^2 + y_0^2}{2|\vec{r}|} - (\alpha \cdot x_0 + \beta \cdot y_0) - \frac{(\alpha \cdot x_0 + \beta \cdot y_0)^2}{2|\vec{r}|} \qquad (2-12)$$

Aufgrund des verhältnismäßig sehr großen Abstands $|\vec{r}|$ können die quadratischen Terme in (2-12) vernachlässigt werden:

$$|\vec{r} - \vec{r}_0| \cong |\vec{r}| - (\alpha \cdot x_0 + \beta \cdot y_0) \qquad (2-13)$$

Dieser Abstand bedeutet in der e-Funktion, dass die Phase der Elementarwellen im Punkt P nur noch linear von x_0 und y_0 abhängig ist. Diese Interpretation stimmt mit der Einteilung der Beugungseffekte überein, die bei [Lipson] vorgenommen wird

²⁸ vgl. [Lipson], S. 5

²⁹ vgl. [Pedrotti], S. 471

(siehe 2.2.2). Diese FRAUNHOFER-Näherung führt zum FRAUNHOFER-Beugungsintegral³⁰:

$$E(\alpha, \beta, z) = -i \frac{n}{\lambda \cdot |\vec{r}|} \cdot \frac{z}{|\vec{r}|} \quad (2-14)$$

$$\cdot e^{ik|\vec{r}|} \iint_A E_0(x_0, y_0, 0) \cdot e^{-2\pi i \frac{n}{\lambda} (\alpha \cdot x_0 + \beta \cdot y_0)} dx_0 dy_0$$

Gerade bei der numerischen Berechnung dieses Integrals sind absolute Koordinaten auf der Zielebene besser zu handhaben. Es stellt keine Schwierigkeiten dar, die Richtungskosinus α und β dementsprechend umzurechnen.

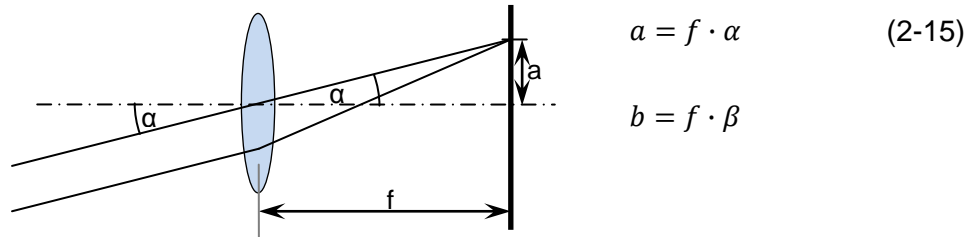


Abbildung 9 Bildentstehung bei FRAUNHOFER-Beugung

Abbildung 9 zeigt Grundlage und Ergebnis einer solchen Koordinatenumwandlung. Es ist dabei zu beachten, dass dieses Koordinatensystem von der z-Position der Zielebene abhängt. Zusammen mit den für die FRESNEL-Näherung erläuterten Vereinfachungen lässt sich (2-14) alternativ darstellen als:

$$E(a, b, z) = -i \frac{n}{\lambda \cdot z} \cdot e^{2\pi i \frac{n}{\lambda} z} \quad (2-16)$$

$$\cdot e^{\pi i \frac{n \cdot z}{\lambda \cdot f^2} (a^2 + b^2)} \iint_A E_0(x_0, y_0, 0)$$

$$\cdot e^{-2\pi i \frac{n}{\lambda \cdot f} (a \cdot x_0 + b \cdot y_0)} dx_0 dy_0$$

Analog zu den Überlegungen zur FRESNEL-Beugung sind auch für die Fernfeld-Beugung Kriterien für die Gültigkeit der genannten Berechnungsformeln festzulegen. Die vernachlässigten quadratischen Terme dürfen in der Phase der Elementarwellen nur einen begrenzten Fehler verursachen. Einen entsprechenden Fehlerfaktor definiert [Träger] ähnlich wie für die FRESNEL-Näherung:

$$Q_{\text{Fraunhofer}} = \frac{n \cdot D^2}{4 \cdot \lambda \cdot z} \ll 1 \quad (2-17)$$

Dieser Faktor erlaubt eine Einschätzung der Genauigkeit der FRAUNHOFER-Näherung in einem untersuchten Bereich. Er stimmt mit der in (2-9) beschriebenen

³⁰ vgl. [Träger], S. 129

FRESNEL-Zahl überein, die in manchen Literaturstellen für die Einteilung der Beugungsgebiete herangezogen wird.

Ein Beispiel mit $n = 1$, $\lambda = 532 \text{ nm}$ und $D = 50 \text{ }\mu\text{m}$ illustriert die Bedeutung der FRAUNHOFER-Beugung für die DOE-Berechnung. Ausgehend von den gegebenen Parametern, wird z abgeschätzt zu:

$$z \gg 3,125 \text{ mm}$$

Abgesehen von diversen Miniaturisierungsanwendungen liegen übliche Arbeitsabstände für DOEs im Bereich mehrerer cm bis dm. Das Ergebnis zeigt, dass bei der Entwicklung die Fernfeld-Beugung eine tragende Rolle spielt.

Darüber hinaus ermöglicht die FRAUNHOFER-Näherung eine Berechnung der komplexen Amplitude im Fokus einer Linse, was ihre Bedeutung unterstreicht³¹.

2.3 FOURIER-Optik

2.3.1 Grundlagen der FOURIER-Theorie

In einer Arbeit zur Wärmeleitung in festen Körpern führte der französische Mathematiker Jean B. J. FOURIER Anfang des 19. Jahrhunderts eine Methode ein, die es erlaubt, eine beliebige periodische Funktion als unendliche Reihe von Sinusfunktionen darzustellen. Die Frequenz jeder Sinusfunktion ist ein ganzzahliges Vielfaches einer Grundfrequenz ω_0 . Jede Funktion ist durch eine Amplitude A_n und eine Phasenlage φ_n definiert:

$$f(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cdot \sin(n\omega t + \varphi_n) \quad (2-18)$$

Diese als FOURIER-Reihen bezeichneten Summen lassen sich auch in den folgenden Formen darstellen:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cdot \cos(n\omega t) + b_n \cdot \sin(n\omega t)) \quad (2-19)$$

$$f(t) = \sum_{n=-\infty}^{\infty} (C_n \cdot e^{in\omega t}) \quad (2-20)$$

³¹ Siehe [Träger], S. 129-132

Besonders die komplexe EULERSche Exponentialdarstellung (2-20) ist aufgrund ihrer mathematischen Vorteile bei Integration und Differentiation sehr gebräuchlich.

Von großem Interesse in vielen technischen und physikalischen Bereichen sind dabei die komplexen Amplituden C_n der einzelnen Frequenzen $n \cdot \omega$. Die Berechnung dieser Koeffizienten wird als FOURIER-Analyse bezeichnet:

$$C_n = \frac{1}{T} \cdot \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cdot e^{-in\omega t} dt \quad (2-21)$$

Auf die mathematischen Hintergründe der FOURIER-Analyse wird in dieser Arbeit nicht eingegangen. Auch zur Umrechnung der entsprechenden FOURIER-Koeffizienten zwischen den unterschiedlichen Schreibweisen sei auf die einschlägige Literatur verwiesen. Die dargestellten Formeln (2-18) bis (2-21) gelten allerdings nur für periodische Funktionen mit der Periodendauer

$$T = \frac{2\pi}{\omega} \quad (2-22)$$

bzw. für nicht periodische Funktionen in einem endlichen Intervall T .

Bei absolut unperiodischen Funktionen läuft T gegen unendlich, was infinitesimal kleine Kreisfrequenzen ω zur Folge hat. Die Summe der FOURIER-Reihe geht in ein Integral über. Aus den diskreten Koeffizienten C_n wird eine stetige Funktion $F(\omega)$, deren Bestimmung ebenso bedeutend ist:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cdot e^{-i \cdot \omega t} dt \quad (2-23)$$

Mathematisch entspricht dies einer Funktionen-Transformation aus dem Zeit- in den Frequenzraum. Obgleich FOURIER selbst diese Erweiterung seiner Theorie nicht gesehen hatte, ist sie als FOURIER-Transformation von $f(t)$ bekannt. Die entsprechende Analogie zur FOURIER-Reihe wird als inverse FOURIER-Transformation bezeichnet:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \cdot e^{i \cdot \omega t} d\omega \quad (2-24)$$

2.3.2 Verknüpfung von FOURIER-Theorie und Beugung

Der mathematische Ausdruck (2-23) für die FOURIER-Transformation entspricht der kontinuierlichen Analyse der Frequenzanteile einer Zeitfunktion $f(t)$. Sie gibt an, mit

welcher Amplitude und Phase die jeweiligen harmonischen Funktionen der Frequenz $t \cdot \omega$ überlagert werden müssen, um $f(t)$ zu erhalten. Zur Berechnung der konkreten komplexen Amplitude einer bestimmten Frequenz werden also alle Funktionswerte von $f(t)$ unter Berücksichtigung ihrer Phase $e^{-i \cdot \omega t}$ aufsummiert.

Die Übertragung dieses Zusammenhangs von einer zeitlichen Funktion $f(t)$ zu einer räumlich abhängigen Größe $f(\vec{r})$ bildet die Grundlage der FOURIER-Optik. Analog werden zur Berechnung der komplexen Amplitude einer bestimmten Raumfrequenz k alle Funktionswerte von $f(\vec{r})$ aufsummiert, wobei die zur jeweiligen Position zugehörige Phasenlage multiplikativ einbezogen wird. Dies ist nichts anderes als eine Beschreibung des HUYGENSSchen Prinzips der Wellenausbreitung.

Die Größe, die summiert (oder genauer: integriert) wird, ist die Feldstärken-Amplitude der Elementarwellen. Ein Vergleich von (2-23) mit dem FRESNEL-Beugungsintegral (2-8) zeigt die FOURIER-Darstellung auch in den Gleichungen. Wird der Term vor dem Integral zu einer Konstanten

$$C(\vec{r}) = -i \frac{n}{\lambda \cdot z} \cdot e^{i \frac{2\pi n \cdot z}{\lambda}} e^{i \pi n \frac{x^2 + y^2}{\lambda \cdot z}} \quad (2-25)$$

und werden im Integral weitere Terme zusammengefasst

$$F(\vec{r}_0) = E_0(\vec{r}_0) \cdot e^{i \pi n \frac{x_0^2 + y_0^2}{\lambda \cdot z}} \quad (2-26)$$

ergibt sich das FRESNEL-Beugungsintegral zu:

$$E(\vec{r}) = C(\vec{r}) \iint_A F(\vec{r}_0) \cdot e^{-i \frac{k}{z} \cdot \vec{r}_0 \cdot \vec{r}} dx_0 dy_0 \quad (2-27)$$

Die Ortskreisfrequenz k ist als analoge Größe zur Kreisfrequenz des Zeitraumes zu verstehen.

$$k = \frac{2\pi}{\lambda} \quad \omega = \frac{2\pi}{T} \quad (2-28)$$

$$\vec{k} = \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \frac{2\pi}{\lambda \cdot z} \cdot \begin{pmatrix} x \\ y \end{pmatrix}$$

Aufgrund des Flächenintegrals müssen \vec{k} und \vec{r} als vektorielle Größen betrachtet werden. $\vec{k} \cdot \vec{r}$ ist das Skalarprodukt beider Vektoren. Die z-Komponente dieser Vektoren ist unter Berücksichtigung der erwähnten Vereinfachungen bereits vor dem Integral in die Konstante $C(\vec{r})$ eingeflossen. Das FRESNEL-Integral lautet damit:

$$E(\vec{k}) = C(\vec{k}) \iint_A F(\vec{r}_0) \cdot e^{-i \vec{k} \cdot \vec{r}_0} dx_0 dy_0 = C(\vec{k}) \cdot \mathcal{F}(F(\vec{r}_0)) \quad (2-29)$$

In gleicher Weise kann auch das FRAUNHOFER-Beugungsintegral verkürzt dargestellt werden:

$$\begin{aligned} E(a, b, z) &= C(a, b, z) \iint_A E_0(\vec{r}_0) \cdot e^{-i\frac{k}{f}(a \cdot x_0 + b \cdot y_0)} dx_0 dy_0 \\ &= C(a, b, z) \cdot \mathcal{F}(E(\vec{r}_0)) \end{aligned} \quad (2-30)$$

Wie (2-30) zu entnehmen ist, kann die Beugungsverteilung im Fernfeld direkt aus der Amplitude in der Beugungsebene über FOURIER-Transformation gewonnen werden. Die Bedingungen für die Gültigkeit der jeweiligen Beugungs-Modelle bleiben unberührt. Der Vorteil einer solchen Darstellung liegt in der Möglichkeit der besonders schnellen Berechnung durch Nutzung der Algorithmen der schnellen FOURIER-Transformation (siehe 2.5.3). Ein weiterer Vorteil dieser kompakten Beschreibung von Beugungs-Phänomenen wird in Abschnitt 2.3.4 erläutert.

In der FRESNEL-Formulierung (2-29) ist die Feldstärkeverteilung in der Bildebene in Abhängigkeit von der Ortskreisfrequenz k angegeben. Damit wird deutlich, dass sich das Beugungsbild nicht mehr im normalen Ortsraum, sondern im reziproken Frequenzraum befindet. Während im Ortsraum die Koordinaten x_0 und y_0 in [m] angegeben werden, sind die Einheiten der Koordinaten k_x und k_y im Frequenzraum [m^{-1}]. Da sich das Bild jedoch physisch auf einem Schirm mit den Koordinaten x und y abbildet, ist es nötig gemäß Gleichung (2-28) die Koordinaten zu überführen.

Entsprechend der Beugungstheorie der Bildentstehung bestimmt die höchste Beugungsordnung, die von einer Linse noch erfasst wird, wie klein die minimal abzubildenden Strukturen sind³². Hohe Beugungsordnungen sind also für kleine Strukturen und feine Details notwendig. Dies lässt sich alternativ auch mit der FOURIER-Theorie erklären.

In der vereinfachten Betrachtung als FOURIER-Reihe kann eine Funktion umso genauer durch die FOURIER-Reihe beschrieben werden, je mehr Reihenglieder einbezogen werden. Wird die Reihe nach den ersten aufsummierten Wellenfunktionen – d.h. nach den Wellen mit den niedrigsten Frequenzen – abgebrochen, bleibt die resultierende Funktion sehr grob mit schlechten Kontrasten und keinen Details. Erst mit Hinzunahme höherer Frequenzen nimmt die Genauigkeit der erzeugten Funktion zu. Angewendet auf die optische Beugung heißt das, dass das dargestellte Beugungsbild nicht nur eine optische Erscheinung ist. Es zeigt vielmehr das Amplitudenspektrum, das zur Abbildung der beugenden Struktur nötig ist. Die aktive Manipulation dieses Spektrums, z.B. über Blenden, wird u.a. in der digitalen Bildbearbeitung oder der optischen Filterung genutzt.

³² vgl. [Lipson], S. 314

2.3.3 Ausgewählte Beispiele

Die erläuterten Zusammenhänge werden anhand zweier Beispiele illustriert. Eine der am häufigsten auftretenden Funktionen ist die Rechteckfunktion und ihre Transformierte. Im Rahmen einer räumlichen Feldstärkeverteilung E_0 entspricht eine Rechteckfunktion einer Blende, die konstant beleuchtet wird. Die Pulsbreite des Rechteckpulses ist in diesem Fall die Apertur der Blende.

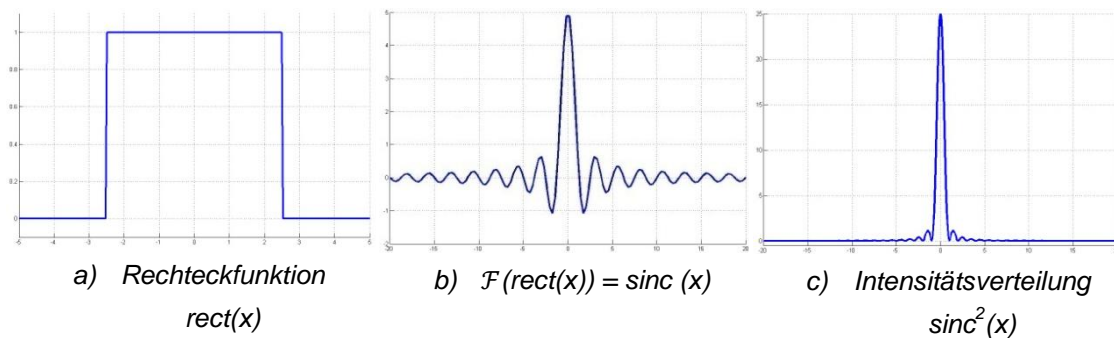
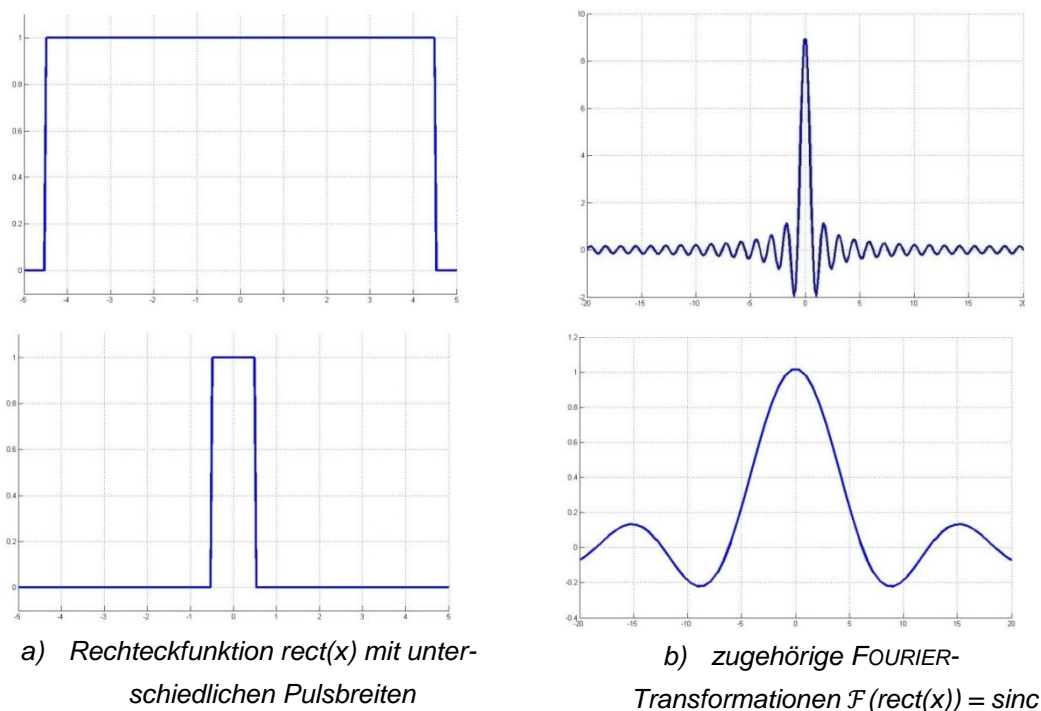


Abbildung 10 Rechteckfunktion und ihre FOURIER-Transformierte (erstellt mit MATLAB)

Abbildung 10 zeigt die Funktion und das Ergebnis ihrer FOURIER-Transformation. Die typische hyperbelförmig abfallende Sinusfunktion wird als *sinus cardinalis* (sinc) bezeichnet. Für die Deutung als Beugungsverteilung ist zu beachten, dass nur Feldstärken aufgetragen sind, die dementsprechend noch zu quadrieren sind. Das Resultat ist der bekannte $\text{sinc}^2(x)$ -Verlauf, der das Beugungsbild an einem Spalt repräsentiert.



(x)

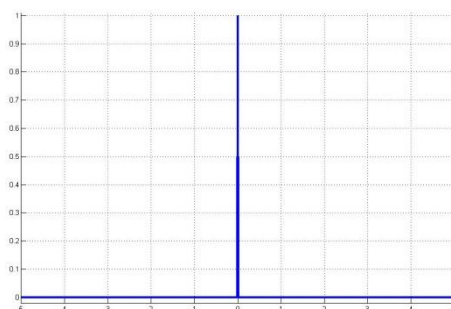
Abbildung 11 Variation der Spaltbreite der Rechteckfunktion
(erstellt mit MATLAB)

Eine Variation der Spalt-/Pulsbreite illustriert die bereits erläuterten reziproken Zusammenhänge. Wie in Abbildung 11 dargestellt ist, führt eine größere Spaltbreite zu einem schmaleren Beugungsbild. Eine Verringerung dagegen resultiert in einer Verbreiterung der sinc-Funktion. Die Spaltbreite und z.B. der Abstand der ersten Extrempunkte in der Transformierten verhalten sich genau reziprok zueinander. Die Richtigkeit der dargestellten Effekte lässt sich auch durch Beugungsexperimente am Spalt praktisch nachprüfen.

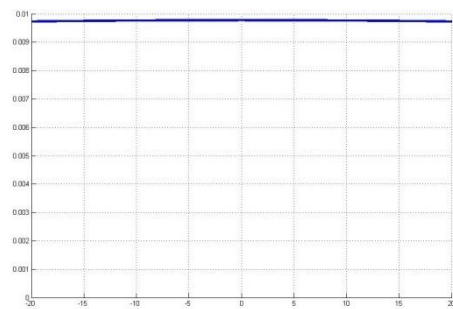
In Übereinstimmung mit den Erklärungen des letzten Abschnitts muss eine schmalere Spaltbreite auch zu jener Verbreiterung des Beugungsbildes führen. Da eine schmalere Rechteckverteilung eine vergleichsweise feinere Struktur darstellt, sind die höheren Frequenzanteile bedeutender für die Generierung des Pulses. Ihre Amplitude steigt deutlich an.

Auch im Bereich der Lasertechnik ist diese Tatsache bekannt. Hier wird zur Erzeugung zeitlich ultrakurzer Laserpulse eine Vielzahl von Wellenlängen benötigt.

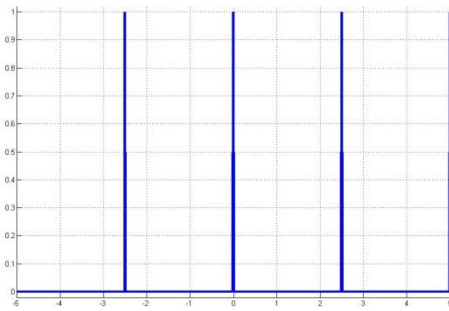
Eine Verringerung der Spaltbreite bis zum Grenzwert 0 führt zur DIRACschen Deltafunktion. Diese ist definiert als ein unendlich schmaler Puls mit unendlich hohem Funktionswert. Die Fläche unter der Funktion ist mit 1 normiert. Abbildung 12 zeigt die Funktion, eine Reihe aus äquidistanten Deltafunktionen und die zugehörigen FOURIER-Transformierten:



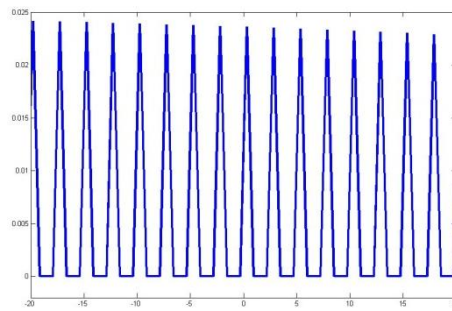
a) Einfache Deltafunktion $\delta(x)$



b) zugehörige FOURIER-Transformation
 $\mathcal{F}(\delta(x)) = \text{konst.}$



c) Reihe von äquidistanten Deltafunktionen $\delta_i(x)$



d) zugehörige FOURIER-Transformation

Abbildung 12 DIRACsche Deltafunktion und ihre Eigenschaften
(erstellt mit MATLAB)

Das Resultat in Bild b) ergibt sich auch aus der Überlegung des Grenzwertes für einen unendlich schmalen Rechteckpuls. Die Deltafunktion ist generell für die Numerik von großer Bedeutung, da das Produkt einer beliebigen Funktion und einer Deltafunktion an einer Stelle x_0 stets den Funktionswert der Funktion ergeben wird. Erweitert zu einer Reihe äquidistanter Deltafunktionen entsteht die mathematische Grundlage für die numerisch diskrete Abtastung kontinuierlicher Funktionen.

Die Eigenschaft, dass eine Reihe von Deltafunktionen mit FOURIER-Transformation erneut zu einer Reihe von Deltafunktionen führt, wird im anschließenden Abschnitt bei der Betrachtung der optischen Faltung eine Rolle spielen. Die sichtbaren Anstiege in Bild d) sind durch die begrenzte numerische Auflösung bedingt.

2.3.4 Mathematische Erweiterungen

Im Folgenden werden noch einige mathematische Ergänzungen der FOURIER-Theorie erklärt, die das Verständnis mancher optischer Phänomene erheblich vereinfachen. Die bereits angedeutete mathematische Operation der Faltung zweier Funktionen hat in der Behandlung von Beugungs-Effekten einen immens wichtigen Platz.

Mathematisch ist die Faltung einer Funktion $f(x)$ mit einer Funktion $g(x)$ definiert durch³³:

$$h(x) = f(x) * g(x) = \int_{-\infty}^{\infty} f(x') \cdot g(x - x') dx' \quad (2-31)$$

³³ vgl. [Lipson], S. 88

Die praktische Vorstellung der Wirkung dieser Operation ist nicht ganz einfach. Am verständlichsten ist möglicherweise die, eine Faltung als Prüfung des größten Überlapps der beiden Funktionen zu betrachten. Die Funktion g wird von $-\infty$ bis $+\infty$ über f geschoben und über das Produkt beider Funktionen integriert. Das ermittelte Ergebnis ist ein Maß für die gemeinsame Fläche unter beiden Funktionen und bildet den Funktionswert der Faltung $h(x)$. Das Maximum der Faltung zeigt den x -Wert der größten Übereinstimmung. Dies wird u.a. im Bereich der optischen Autokorrelation eingesetzt³⁴.

Da eine Deltafunktion als Produkt mit einer Funktion g stets deren Funktionswert liefert, wird eine Faltung aus beiden Funktionen stets die Funktion g ergeben. Eine Reihe von Deltafunktionen ergibt dementsprechend eine Reihe Funktionen g , deren Referenzpunkt sich jeweils an den Orten der diskreten Deltafunktionen befindet.

Die Faltung zweier Funktionen f und g bekommt im Zusammenhang mit dem Faltungs-Theorem eine weitere Bedeutung. Sinngemäß besagt dieses Theorem, dass eine Faltung im Ortsraum und eine Multiplikation im FOURIER-Raum ein sogenanntes FOURIER-Transformationspaar bilden:

$$f(x) * g(x) = F(k) \cdot G(k) \qquad f(x) \cdot g(x) = F(k) * G(k) \qquad (2-32)$$

Das Faltungs-Theorem lässt sich auch in die andere Richtung anwenden. Das heißt, einer Faltung im FOURIER-Raum steht immer ein Produkt im Ortsraum gegenüber. $F(k)$ und $G(k)$ sind die jeweiligen FOURIER-Transformierten der Funktionen $f(x)$ und $g(x)$.

Auch das FRESNEL-Beugungsintegral (2-8) lässt sich als Faltung zweier Funktionen ausdrücken³⁵. Dieser Ausdruck beschreibt das Beugungsintegral im FOURIER-Raum und bietet verschiedene Vorteile bei der Berechnung.

Die Berechnung komplizierter Beugungsmuster lässt sich durch Anwendung des Faltungstheorems erheblich vereinfachen. Es ist damit möglich, die Beugungsmuster von Strukturen, die sich als Faltung interpretieren lassen, durch einfache Multiplikation ihrer elementaren Beugungsbilder zu bekommen. Und umgekehrt erleichtert das Theorem auch die Ermittlung der Beugungsverteilung von überlagerten Aperturen, da sich diese als Multiplikation in der Beugungsebene darstellen lassen.

Ein binäres Transmissionsgitter lässt sich auf diesem Weg z.B. als Faltung einer unendlichen Reihe von Deltafunktionen mit einem einfachen Rechteckspalt beschreiben. Die einzelnen Deltafunktionen der Reihe befinden sich jeweils an den Orten der Gitteröffnungen und der Rechteckspalt beinhaltet die spezifischen Gitter-

³⁴ vgl. [Pedrotti], S. 748-752

³⁵ vgl. [Träger], S. 128

parameter. Gemäß dem Faltungstheorem ergibt diese Anordnung im Beugungsbild eine Multiplikation von sinc- und Deltafunktion bzw. dem Quadrat daraus. Das Ergebnis ist gewissermaßen eine abgetastete sinc-Funktion.

Eine zweite mathematische Erweiterung, die im Rahmen dieser Arbeit weder hergeleitet noch weitgehend analysiert werden soll, ist das PARSEVALSche Theorem. Mathematisch ausgedrückt, lautet es:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = C \cdot \int_{-\infty}^{\infty} |F(k)|^2 dk \quad (2-33)$$

Der Inhalt des Theorems lässt sich physikalisch einfach als Energieerhaltungssatz interpretieren. Es sagt aus, dass die Energie der Aperturverteilung gleich oder zumindest proportional zur Energie der Beugungsverteilung sein muss.³⁶

2.4 Design diffraktiver optischer Elemente

2.4.1 Grundsätzliches

Im letzten Abschnitt wurde mit der FOURIER-Optik ein elementares Berechnungsprinzip eingeführt, das es besonders komfortabel erlaubt, für eine beliebige komplexe Verteilung in der Aperturebene die resultierende Beugungsverteilung in der Bildebene zu berechnen. Gerade die fortgeschrittenen rechentechnischen Lösungsmöglichkeiten der FOURIER-Transformation (siehe 2.5.3) mittels numerischer, computerbasierter Verfahren eröffnen neue Aussichten für das Optikdesign.

In diesem Zusammenhang ist der umgekehrte Weg von großem Interesse: Wie ist es möglich, für eine vorgegebene Verteilung in der Bildebene die notwendige komplexe Transmissionsfunktion in der Aperturebene zu ermitteln?

Es zeigt sich, dass keine ideale Lösung existiert, die für alle denkbaren verbundenen Problemstellungen ausreichend genau und gleichzeitig effizient funktioniert. Bereits in den vorangegangenen Abschnitten ist auf die verschiedenen Varianten des Beugungsintegrals eingegangen worden. Es ist vom Abstand zwischen Apertur und Bildebene abhängig, ob FRESNEL-, FRAUNHOFER- oder KIRCHHOFF-Integral genutzt werden kann. Zudem muss bei der Ermittlung einer geeigneten Transmissionsfunktion der optische Aufbau, für den das DOE vorgesehen ist, beachtet werden. Dieser beeinflusst das Ergebnis erheblich.

³⁶ vgl. [Lipson], S. 94

Die häufigsten optischen Anordnungen sind die 1f- und die 2f-Anordnung und die Winkelabstrahlung. Daneben existieren noch Fernfeld- und FRESNEL-Aufbau. Welche der zum Teil in Abbildung 13 illustrierten Arrangements zum Einsatz kommt, ist von der anschließenden Anwendung abhängig.

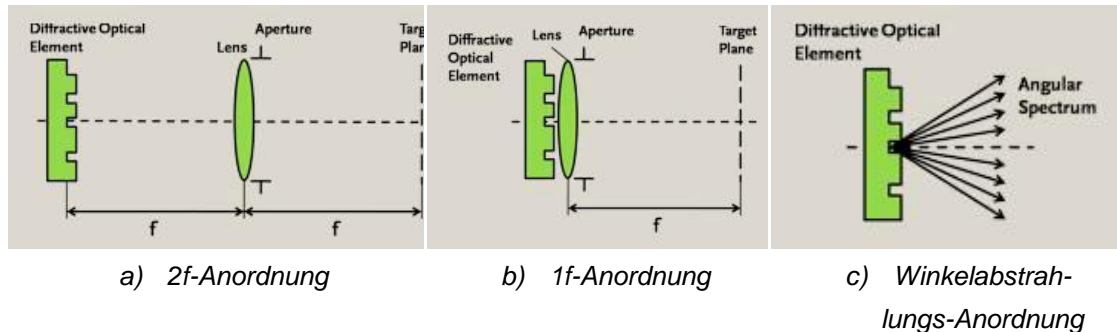


Abbildung 13 Varianten von DOE-Anordnungen
[aus Software VirtualLab 4 entnommen]

Neben den nachfolgend vorgestellten Verfahren und Algorithmen zur Berechnung von DOEs gibt es noch eine Vielzahl andere. Der Ansatz der sogenannten Multi Grid Arrays formt die einfallende Wellenfront durch Anordnung kleiner Phasengitterelemente in einem Array. Auf jedes dieser Mini-Gitter fällt ein Anteil der einfallenden Welle und das jeweils typische Gitter-Beugungsbild wird erzeugt. Durch geschickte Anordnung der Einzelgitter und Variation von Gitterperiode und –art können vielfältige Ausgangsverteilungen erzeugt werden. Die bereits erwähnte Software VirtualLab nutzt diesen Ansatz in einer Toolbox zur Erzeugung individueller Beleuchtungsprofile.

Ein weiteres Verfahren, das hier am Rande erwähnt werden soll, ist das Simulated Annealing. Dies ist ein globaler Optimierungsalgorithmus, der seinen Ursprung in einem physikalischen Modell der langsamen Abkühlung eines Systems hat. Analog zum sich dabei einstellenden stabilsten Zustand niedrigster thermischer Energie wird beim Simulated Annealing eine charakteristische Gewinnfunktion minimiert, deren Minimum das optimale Ergebnis einer Aufgabe darstellt. Die Stärke dieses Verfahrens liegt in eingebauten Sprüngen, durch die der Algorithmus seltener in einem lokalen Minimum „stecken bleibt“. Somit kann treffsicherer ein globales Minimum gefunden werden. Im Beispiel der DOE-Berechnung werden so die DOE-Parameter solange variiert, bis das Ergebnis der gewünschten Feldverteilung am ähnlichsten ist.³⁷

³⁷ vgl. [Stankovic], S. 94 - 98

2.4.2 Analytische Verfahren

Nach der Theorie aus 2.3.2 müsste eine mathematisch einfache Lösung dieses Problems die Anwendung der FOURIER-Optik sein. Im optischen Aufbau, wie er in Abbildung 13a dargestellt ist, wird das Beugungsbild des DOEs durch die Linse in der Bildebene abgebildet. Gemäß (2-30) ist dieses Beugungsbild die FOURIER-Transformierte der komplexen Feldverteilung der DOE-Ebene. Umgekehrt ergibt die inverse FOURIER-Transformation eines gewünschten Beugungsbildes die komplexe Transmissionsfunktion des verwendenden DOEs.

Mathematisch betrachtet ist dieser Gedankengang richtig und tatsächlich werden ähnliche Verfahren im Bereich der digitalen Bildbearbeitung seit langem mit Erfolg eingesetzt. Das physikalische Problem, welches jedoch in dieser Betrachtung vernachlässigt wurde, ist, dass in der Bildebene lediglich die Intensität gemessen wird. Wird also eine spezielle gewünschte Intensitätsverteilung in der Bildebene vorgegeben, kann lediglich eine Aussage über die Amplitude des Feldes getroffen werden. Die Information der Phasenlage der Welle fehlt.

Wie in den nachfolgenden Unterabschnitten gezeigt werden wird, ist die inverse FOURIER-Transformation nichtsdestotrotz als wichtiges Werkzeug Teil des Lösungsprozesses bei der Realisierung der verschiedenen Algorithmen.

Des Weiteren existieren noch andere Möglichkeiten, ein DOE auf analytischem Weg zu berechnen. Durch Anwendung diverser Ray-Tracing-Verfahren sind auch Lösungen für einfache, analytisch beschreibbare Anforderungen möglich. Dazu gehört z.B. die Umwandlung eines Gaußstrahls in einen rechteck- oder kreisförmigen Top-Hat-Strahl³⁸. Da diese Verfahren aber keine Lösungsmöglichkeit für allgemeine Probleme anbieten, soll deren Erwähnung in dieser Arbeit genügen.

2.4.3 LOHMANN-Methode

Die Methode von LOHMANN, auch „Phasen Umleitungs Methode“ (detour phase method) genannt, ist bereits seit über 40 Jahren bekannt und eng mit dem Begriff Computergeneriertes Hologramm verbunden. Als einfach zu verstehende und programmiertechnisch relativ leicht zu realisierende Methode wird das Grundprinzip erklärt und auf mögliche Schwächen eingegangen.

Die mit der LOHMANN-Methode erzeugten Elemente sind reine binäre Amplituden-DOEs und daher auch praktisch verhältnismäßig leicht herzustellen. Sie sind jeweils für einen optischen 2f-Aufbau ausgelegt (siehe Abbildung 13). Es wird von den Vo-

³⁸ siehe [Guessous], S. 46-52

raussetzungen für FRAUNHOFER-Beugung ausgegangen, d.h. die Phasenfronten der Wellen in der Beugungsebene sind planar.

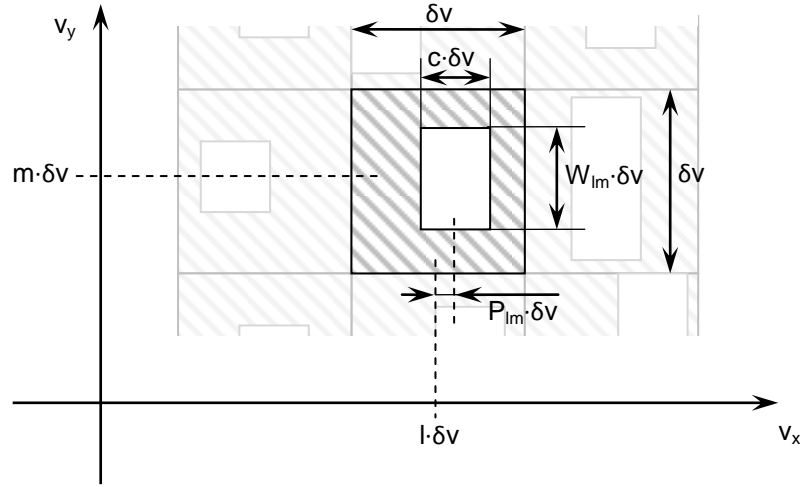


Abbildung 14 Prinzip der DOE-Diskretisierung in der LOHMANN-Methode

Die Beugungsebene wird hierbei in kleine Quadrate diskretisiert, die alle eine individuelle rechteckige Apertur beinhalten. Die Größe dieser Blende steuert für jeden Rasterpunkt die transmittierte Amplitude über die Parameter W_{lm} und c . W_{lm} muss für jede Zelle ermittelt werden, während c global optimiert werden muss. Die Position der Apertur steuert die Phasenlage der Einzelwelle. Dies erfolgt mittels des Parameters P_{lm} , der die horizontale Verschiebung steuert. In vertikaler Richtung sind die Zellen zentriert. Abbildung 14 verdeutlicht die prinzipielle Anordnung der Apertur in einer Elementarzelle.

Die komplexe Transmissionsfunktion $f(v_x, v_y)$ des DOEs wird vollständig real, da in z -Richtung nur eine Modulation der Amplitude erfolgt. Sie ist 1 in den Bereichen der Elementarblenden und 0 an allen anderen Stellen, oder mathematisch ausgedrückt:

$$f(v_x, v_y) = \sum_l \sum_m \text{rect} \left[\frac{v_x - (l + P_{lm}) \cdot \delta v}{c \cdot \delta v} \right] \cdot \text{rect} \left[\frac{v_y - m \cdot \delta v}{W_{lm} \cdot \delta v} \right] \quad (2-34)$$

Die Rechteckfunktion rect beschreibt die Durchlässigkeit der Blende und springt jeweils zwischen 0 und 1. Die Transmissionsfunktion ist im beschriebenen Fall nicht mehr von den Ortskoordinaten auf der Beugungsebene abhängig, sondern von den Ortsfrequenzen v_x und v_y . Diese Koordinatentransformation ist notwendig, da ansonsten das gewünschte Beugungsbild nach der FOURIER-Transformation im Ortsfrequenz-Raum angegeben wird (vgl. 2.3.2).

Eine planare, über die y -Achse verkippt einfallende Welle kann wie folgt beschrieben werden:

$$E_0(v_x) \sim e^{i2\pi \cdot x_0 \cdot v_x} \quad (2-35)$$

Der Grund für die Annahme einer Verkipfung der Eingangswelle geht aus der verwendeten Quelle³⁹ nicht klar hervor und wird auch nicht weiter erläutert. Eine Vermutung ist, dass diese allgemeinere Formulierung die Funktion des DOEs auch bei nicht exakt senkrechtem Lichteinfall sicher stellt. Der Sonderfall exakt zur Blende paralleler Wellenfronten ist darin mit $x_0 = 0$ enthalten.

Die sich in der Beugungsebene ergebende komplexe Amplitude

$$E_0(v_x, v_y) = E_0 \cdot f(v_x, v_y) \cdot e^{i2\pi \cdot x_0 \cdot v_x} \quad (2-36)$$

erzeugt in der Bildebene das durch ihre FOURIER-Transformation beschriebene Beugungsbild:

$$E(x, y) = E_0 \cdot \iint_A f(v_x, v_y) \cdot e^{i2\pi \cdot [(x+x_0) \cdot v_x + y \cdot v_y]} dv_x dv_y \quad (2-37)$$

Das Ergebnis dieser aufwändig zu integrierenden Funktion beträgt gemäß [Ersoy]:

$$\begin{aligned} E(x, y) = E_0 \cdot c(\delta v)^2 \cdot \text{sinc}[c\delta v \cdot (x + x_0)] \\ \cdot \sum_l \sum_m [W_{lm} \cdot \text{sinc}(y \cdot W_{lm} \cdot \delta v) \\ \cdot e^{i2\pi \cdot [\delta v \cdot ((x+x_0) \cdot (l+P_{lm}) + y \cdot m)]}] \end{aligned} \quad (2-38)$$

Die enthaltene Rasterbreite δv wird zu Beginn des Prozesses als Konstante vorgegeben. Sie kann z.B. durch das angestrebte Herstellungsverfahren und den realisierbaren Rechenaufwand bestimmt sein. Es ist zu beachten, dass bei einer diskreten FOURIER-Transformation (DFT, siehe 2.5.3) aus mathematischen Gründen das Abtastintervall im Raumbereich und der Abtastabstand im Frequenzbereich zueinander reziproke Größen sind. In diesem Fall gilt dieser Zusammenhang für die Rasterbreite δv und die maximale Ausdehnung des gewünschten Bildes X bzw. Y. Auch diese Einschränkung von δv muss berücksichtigt werden.

Die Parameter x_0 und c müssen global für die gesamte Blendenregion optimiert werden, d.h. sämtliche Elementaraperturen weisen in x-Richtung die gleiche Breite auf. Die Parameter W_{lm} und P_{lm} müssen für jede Elementarzelle so bestimmt werden, dass sich das ergebende Beugungsbild einer gesuchten Verteilung annähert.

Eingehender betrachtet, sind für Gleichung (2-38) noch Vereinfachungen möglich. Der Term $\text{sinc}[c\delta v \cdot (x + x_0)]$ erzeugt über das gesamte erzeugte Bild einen Intensitätsabfall in x-Richtung. Im Allgemeinen erzeugen Rechteckspalte wie die Aperturen des DOEs stets Beugungsbilder, die der sinc-Funktion folgen. Je kleiner die

³⁹ [Ersoy], S. 247

Spaltbreite ist, umso größer wird die Periode dieser Funktion, d.h. umso weiter liegen die höheren Beugungsordnungen auseinander (siehe 2.3.3). Wird die Konstante c also generell klein gewählt, wird der fehlerhafte Einfluss dieses Terms erheblich reduziert und der Intensitätsabfall wird im Bildbereich vernachlässigbar. Allerdings geht dies natürlich auf Kosten der Gesamthelligkeit des Bildes. Es ist hier also ein angemessener Kompromiss zwischen Helligkeit und Homogenität zu finden.

Ein einfacher Ansatz zur Vermeidung dieses Problems ist die Division der gewünschten Bildverteilung durch jenen Term, so dass dem Abfall in x-Richtung schon von Beginn an entgegen gewirkt wird.

Im Folgenden soll angenommen werden, dass c so klein gewählt wird, dass im Bildbereich gelten kann:

$$\text{sinc}[c\delta v \cdot (x + x_0)] \approx 1 \quad (2-39)$$

Der zweite sinc-Term in (2-38) $\text{sinc}(y \cdot W_{lm} \cdot \delta v)$ kann analog zur x-Richtung betrachtet werden. Da W_{lm} jedoch für jede Zelle erst noch individuell zu ermitteln ist, kann eine Reduzierung des Intensitätsabfalles auf dem oben beschriebenen Weg mittels Korrektur des Zielbildes nicht erfolgen. Glücklicherweise ist dieser Fehler nicht so kritisch, wie der in x-Richtung. Wie bereits erklärt wurde, sind X bzw. Y und δv zueinander reziprok. Es gilt also:

$$Y = \frac{1}{\delta v} \quad \text{bzw.} \quad Y \cdot \delta v = 1 \quad (2-40)$$

W_{lm} darf entsprechend der Festlegung (siehe Abbildung 14) nur Werte zwischen 0 und 1 annehmen und y hat als größten Wert $\pm Y/2$. Das Argument des sinc-Terms hat somit die maximale Größe:

$$\begin{aligned} |y \cdot W_{lm} \cdot \delta v| &= \left| \frac{Y}{2} \cdot W_{lm} \cdot \delta v \right| \\ |y \cdot W_{lm} \cdot \delta v| &\leq \frac{1}{2} \end{aligned} \quad (2-41)$$

Gemäß (2-41) beträgt der maximal mögliche Intensitätsabfall im Bildbereich

$$\text{sinc}(0,5) \approx 0,9589$$

und kann somit ebenfalls in ausreichender Näherung mit 1 approximiert werden.

Bereits vor 40 Jahren stellte LOHMANN in seiner Veröffentlichung fest, dass sich Amplitudenfehler deutlich schwächer auf das Bild auswirken als Phasenfehler. Eine Multiplikation von W_{lm} mit einem konstanten Faktor <1 bringt die Möglichkeit, diese Abweichung noch stärker zu unterdrücken, mindert allerdings erneut die Helligkeit des Bildes.

Zuletzt kann die e-Funktion aus (2-38) noch auf folgende Weise zerlegt werden:

$$\begin{aligned} e^{i2\pi \cdot [\delta v \cdot ((x+x_0) \cdot (l+P_{lm}) + y \cdot m)]} \\ = e^{i2\pi \cdot [x \cdot P_{lm} \cdot \delta v]} \cdot e^{i2\pi \cdot [\delta v \cdot (x_0 \cdot (l+P_{lm}) + x \cdot l + y \cdot m)]} \end{aligned} \quad (2-42)$$

Der extrahierte Phasenfaktor verursacht eine Veränderung der Phase der jeweiligen Welle in Abhängigkeit von der x-Position in der Bildebene. Es kann auch hier gezeigt werden⁴⁰, dass diese Abweichung noch innerhalb des $\lambda/4$ -RAYLEIGH-Kriteriums für Wellen-Aberrationen liegt. Auch dieser Teil darf mit annähernd 1 angenommen werden.

Alle Vereinfachungen einbezogen, kann (2-38) geschrieben werden als:

$$E(x, y) = E_0 \cdot c(\delta v)^2 \cdot \sum_l \sum_m [W_{lm} \cdot e^{i2\pi \cdot [\delta v \cdot (x_0 \cdot (l+P_{lm}) + x \cdot l + y \cdot m)]}] \quad (2-43)$$

Wird eine FOURIER-Transformierte wiederum einer inversen FOURIER-Transformation unterzogen, ergibt sich wieder die ursprüngliche Funktion. Dementsprechend kann die gewünschte Feldverteilung $E_{Ziel}(x, y)$ in der Bildebene geschrieben werden als:

$$\begin{aligned} E_{Ziel}(x, y) &= FT^{-1}[FT(E_{Ziel})] \\ &= \iint_A FT(E_{Ziel}) \cdot e^{i2\pi \cdot (x \cdot v_x + y \cdot v_y)} dv_x dv_y \end{aligned} \quad (2-44)$$

Die Numerik erfordert eine diskrete Schreibweise:

$$E_{Ziel}(x, y) = \sum_l \sum_m [F(l\delta v, m\delta v) \cdot e^{i2\pi \cdot [\delta v \cdot (x \cdot l + y \cdot m)]}] \quad (2-45)$$

Der Vergleich von (2-43) und (2-45) zeigt, dass die DFT von E_{Ziel} folgendem Term entspricht:

$$F(l\delta v, m\delta v) = E_0 \cdot c(\delta v)^2 \cdot W_{lm} \cdot e^{i2\pi \cdot [\delta v \cdot x_0 \cdot (l+P_{lm})]} \quad (2-46)$$

Die zweidimensionale FOURIER-Transformation von E führt im Ergebnis zu einem diskreten Raster, in dem jedem Punkt eine Amplitude A_{lm} und eine Phase Φ_{lm} zugewiesen ist.

$$F(l\delta v, m\delta v) = A_{lm} \cdot e^{i2\pi \cdot [\frac{\Phi_{lm}}{2\pi}]} \quad (2-47)$$

⁴⁰ siehe [Ersoy], S. 248

Werden die multiplikativen Konstanten vernachlässigt, ergibt ein Koeffizientenvergleich mit (2-46):

$$W_{lm} = A_{lm} \quad (2-48)$$

$$l + P_{lm} = \frac{\Phi_{lm}}{2\pi \cdot \delta v \cdot x_0}$$

Das Prinzip der LOHMANN-Methode kann somit in folgende fünf Schritte eingeteilt werden:

1. x_0 (Verkippung), δv und die gewünschte Zielfunktion $E_{\text{Ziel}}(x,y)$ festlegen
2. Diskrete FOURIER-Transformation von E_{Ziel} durchführen, das Ergebnis sollte auf den Betrag der Maximalamplitude normiert werden
3. Jedem Punkt des DOE-Rasters die zugehörige Amplitude zuweisen:

$$W_{lm} = A_{lm}$$

4. Jedem Punkt des DOE-Rasters die zugehörige Phase zuweisen:

$$P_{lm} = \frac{\Phi_{lm}}{2\pi \cdot \delta v \cdot x_0} - l$$

5. Konstante c hinsichtlich Helligkeit und Kontrast optimieren.

Die Einfachheit dieses Verfahrens ist einer der großen Vorteile der LOHMANN-Methode. Sie ist sicher historisch bedingt, da die verfügbare Rechentechnik um 1970 keine ausufernden numerischen Algorithmen gestattete. Nur eine durchzuführende FFT ist äußerst sparsam im Vergleich zu anderen iterativen Verfahren. Die erzielbaren Ergebnisse sind trotzdem durchaus zufriedenstellend, wie Abbildung 15 zeigt. Das gewünschte Bild ist in der Rekonstruktion klar erkennbar.

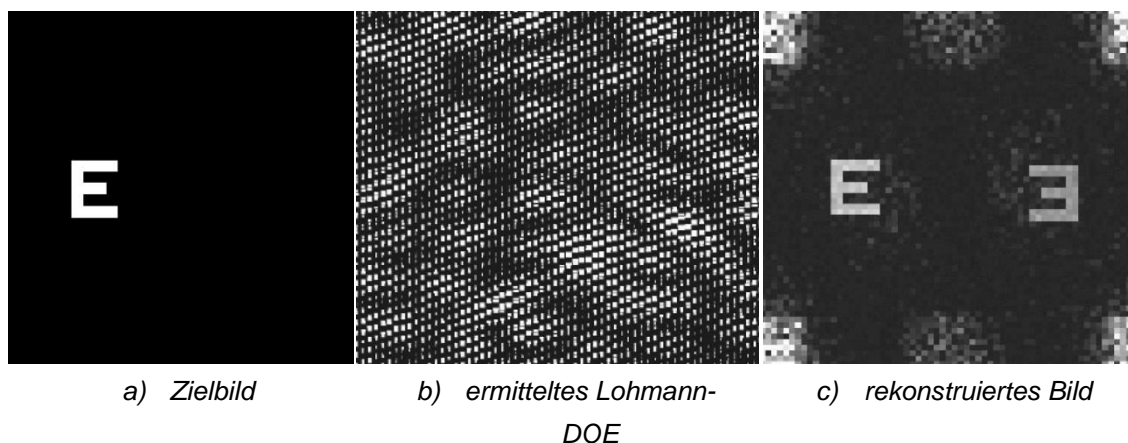


Abbildung 15 Ergebnisse eines mit der Lohmann-Methode berechneten DOE
(Quelle: [Ersoy], S. 251—252)

Das abgebildete DOE wurde mit der Konstanten $c = 0,5$ und in einer weiteren Variante des LOHMANN-Verfahrens erstellt: dem quantisierten LOHMANN-Verfahren. Die Besonderheit besteht darin, dass nicht beliebige Werte für P_{lm} und W_{lm} erlaubt sind,

sondern nur gequantelte diskrete Werte. Im dargestellten Beispiel sind dies 16 verschiedene Stufen.

Die Tatsache, dass nur Amplituden-DOEs erstellt werden können, ist in den meisten Fällen hinderlich, da sie immer mit einem Helligkeitsverlust des Zielbildes einhergeht. Für Wellenlängen, für die keine transmittierenden Materialien vorhanden sind, können diese DOEs allerdings auch eingesetzt werden (z.B. Röntgenstrahlen).

Neben dem auffälligen zweiten Bild, welches jedoch mit weiteren Verbesserungen des Verfahrens auch unterdrückt werden kann, ist das sehr starke Rauschen an den Rändern des Bildbereiches nachteilig für das Verfahren zu bewerten.

2.4.4 Technik iterativer Verschachtelung (IIT und ODIFIIT)⁴¹

Die Technik der iterativen Verschachtelungen (engl. Iterative Interlacing Technique, IIT) ist kein eigentlicher Algorithmus zur Ermittlung einer DOE-Struktur. Dieses Verfahren wird vielmehr auf bestehende Methoden angewandt und ermöglicht ihre Verbesserung hinsichtlich Leistung und Qualität.

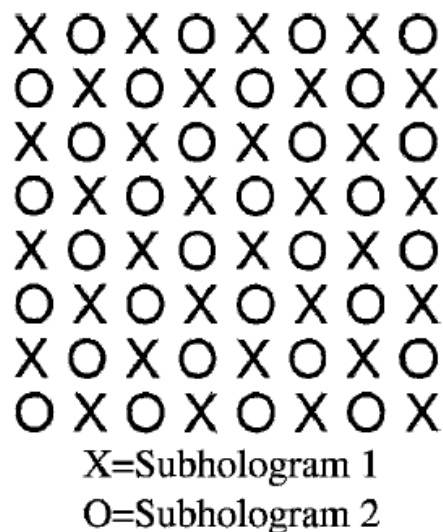


Abbildung 16 Verschachtelung von zwei Unter-DOEs
(Quelle: [Ersoy], S. 290)

Ausgangspunkt ist die Verschachtelungs-Technik (Interlacing Technique, IT), bei der die gesamte DOE-Fläche in eine Menge von Unter-DOEs aufgeteilt wird. Diese Substrukturen können sich auch periodisch wiederholen. Abbildung 16 zeigt eine mögliche Anordnung von zwei verschiedenen Unter-DOEs (Subhologram 1 und 2), es sind aber auch deutlich mehr als zwei DOEs denkbar. Charakteristisch für das

⁴¹ vgl. [Ersoy], S. 289-291

Verfahren ist, dass jedes Unter-DOE separat mittels eines der vorgestellten Methoden designed wird und danach die Substrukturen in der unten dargestellten Weise miteinander verschachtelt werden.

Das erste zu berechnende DOE wird an dem gewünschten perfekten Ausgangsbild $f(m,n)$ orientiert. Sowohl Bild- als auch DOE-Ebene seien in $M \times N$ Pixel diskretisiert. Aufgrund der begrenzten Größe der Substruktur wird die Zielverteilung i.d.R. nicht erreicht. Das rekonstruierte Bild $h_1(m,n)$ weist somit noch einen lokalen Fehler $e_1(m,n)$ auf.

$$e_1(m,n) = f(m,n) - \lambda_1 \cdot h_1(m,n) \quad (2-49)$$

λ_1 ist in (2-49) ein Skalierungsfaktor, der angepasst werden muss. Die Aufgabe des nächsten Teil-DOEs ist es, diesen Bildfehler des ersten DOEs zu minimieren und, wenn möglich, auszulöschen. Das Ziel dieses zweiten DOEs ist somit ein Bild, das ausgedrückt werden kann mit:

$$f(m,n) = \frac{e_1(m,n)}{\lambda_1} \quad (2-50)$$

Im Idealfall erzeugen beide DOEs zusammen dann das gewünschte Bild. Realistisch ist jedoch eine Fortführung dieses Prinzips auf weitere Unter-DOEs, so dass jedes nachfolgende DOE die Bildfehler seiner Vorgänger korrigiert. Sind alle DOEs berechnet, wird das komplett verschachtelte und rekonstruierte Bild $h_1(m,n)$ noch einen finalen Fehler $e_f(m,n)$ aufweisen. Die Minimierung dieses Fehlers ist Ziel der iterativen Erweiterung dieser Technik.

Der Grundgedanke ist die Erneuerung der Teil-DOEs nach dem vollständigen ersten Design. Das neue Wunschbild des ersten Unter-DOEs soll dann den Verlauf haben:

$$f'(m,n) = h_1^1(m,n) - \frac{e_f(m,n)}{\lambda_f} \quad (2-51)$$

mit

$$e_f(m,n) = f(m,n) - \lambda_f \cdot h_f(m,n) \quad (2-52)$$

λ_f ist der zuletzt angewandte Skalierungsfaktor und $h_1^1(m,n)$ die rekonstruierte Feldverteilung nach dem ersten Schritt. Nach dem erfolgten Neudesign des ersten Sub-DOEs wird erneut das Gesamtergebnis ermittelt. Dieses dient als Eingangsgröße zum Neudesign des zweiten Unter-DOEs. Der Fehler kann z.B. als mittlere quadratische Abweichung zwischen rekonstruiertem und gewünschtem Bild definiert werden. Im Zuge dieses iterativen Verfahrens werden die Unter-DOEs so oft nachberechnet, bis der Fehler einen kritischen Wert unterschreitet bzw. sich nicht weiter verringern lässt.

Ein zusätzlicher Vorteil dieses Verfahrens neben der Erhöhung von Leistung und Präzision ist, dass er weniger dazu neigt, in lokalen Minima zu bleiben. Er tendiert meist zum globalen Minimum der Fehlerfunktion oder wenigstens zu einem sehr tiefen lokalen Minimum.

Eine letzte Erweiterung der IIT, die in dieser allerdings Arbeit nicht eingehender behandelt wird, ist die Optimal Decimation-in-Frequency iterative Interlacing Technique (kurz: ODIFIIT). Dieses Prinzip wurde in den 90er Jahren entwickelt, um die Ergebnisse der IIT-Prozedur zusätzlich zu optimieren. Ganz grob gesagt, nutzt dieses Verfahren eine Eigenschaft der FFT-Algorithmen aus, die eine Dämpfung im Frequenzraum verursacht.

Dieses Verfahren reduziert die Dimension, in der die FFT und iFFT ausgeführt werden und verringert so die benötigte Rechendauer. Zudem wird das Design jedes Unter-DOEs effektiver gestaltet⁴².

2.4.5 Iterativer FOURIER-Transformations-Algorithmus (IFTA) ⁴³

Der IFT-Algorithmus stellt einerseits eine spezielle Form der allgemeinen POCS-Methode mit $M = 2$ Randbedingungen dar. Andererseits wird mit diesem Begriff wiederum eine ganze Gruppe von Algorithmen beschrieben. Gemeinsam haben diese alle das Wechseln zwischen verschiedenen Räumen mittels FOURIER-Transformation. Im untersuchten Anwendungsfall sind diese beiden Räume die beugende Aperturebene und die Bildebene.

Den Ursprung des IFTA bildet der Fehlerreduzierungs-Algorithmus von GERCHBERG und SAXTON zur Phasenberechnung computergenerierter Hologramme am Anfang der 1970er Jahre. Am Grundkonzept hat sich bis zum heutigen Tag nichts geändert.

Die beiden gegebenen Randbedingungen, die gewünschte Zielintensitätsverteilung und die Eingangsintensitätsverteilung, wird jeweils zur Korrektur berechneter komplexer Amplituden in Bild- und DOE-Ebene genutzt. Von einer Startamplitude in der Bildebene ausgehend, die sich aus einer zufälligen Phasenlage und der Zielamplitude zusammensetzt, wird über die inverse FOURIER-Transformation die zugehörige komplexe Amplitude in der DOE-Ebene ermittelt. Würde das Ergebnis-DOE nun die berechnete Phasenverschiebung erzeugen und mit der berechneten Amplitudenverteilung bestrahlt, ergäbe sich exakt die gewünschte Verteilung. Hier treten die DOE-Randbedingungen in Kraft.

⁴² Für nähere Informationen, siehe [Ersoy], S. 293-297

⁴³ vgl. [Ripoll]

Hauptsächlich wird hier die Amplitude durch die gegebene Eingangsamplitude korrigiert, d.h. ersetzt. In verschiedenen Varianten des IFT-Algorithmus können an dieser Stelle noch weitere Einschränkungen angewandt werden, wie z.B. die Diskretisierung der Phasenlagen. Im einfachsten Fall wird nun mittels FOURIER-Transformation das Beugungsbild des ggf. angepassten DOEs ermittelt, das sich durch Bestrahlung mit den Eingangsparametern ergibt.

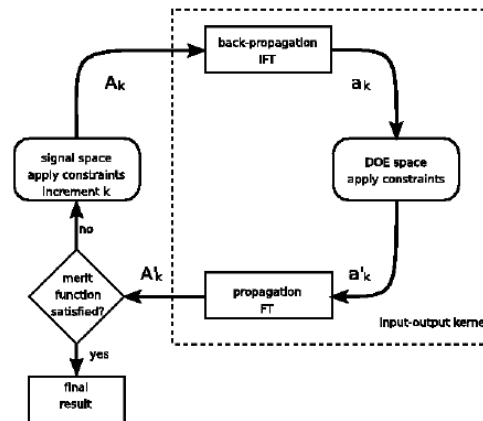


Abbildung 17 Prinzip des IFT-Algorithmus
(Quelle: [Ripoll], S. 2)

Wie Abbildung 17 zeigt, wird das entstandene Beugungsbild nun individuellen Kriterien unterworfen, die beurteilen, wie gut die Annäherung an das Zielbild gelungen ist. Diese Kriterien können rein subjektiv sein oder numerisch berechnet werden. Ein Beispiel für eine solche charakteristische Größe ist die mittlere quadratische Abweichung von Zielintensitätswerten zu den berechneten Intensitätswerten in jedem Punkt der Bildebene.

Im Falle einer negativen Beurteilung wird die Amplitude erneut durch die angestrebte Verteilung korrigiert und der Algorithmus startet von Neuem. Der allgemeine IFT-Algorithmus lässt sich damit wie folgt zusammenfassen:

1. Gewünschte Amplitude mit zufälliger Phase ergibt A_1
2. Inverse FOURIER-Transformation von A_k ergibt a_k
3. Amplitude von a_k wird mit Eingangsamplitude korrigiert und ergibt a'_k
4. FOURIER-Transformation von a'_k ergibt rekonstruiertes Bild A'_k
5. Errechnete Ausgangsamplitude von A'_k mit gewünschter Amplitude korrigieren ergibt A_{k+1}
6. Weiter mit 2.

Dieser überaus simple Algorithmus konvergiert für kontinuierliche Phasenlagen sehr gut und ist einfach nutzbar. Für die praktische Realisierung von DOEs sind kontinuierliche Phasenverteilungen jedoch schwer herzustellen. Werden nur noch diskrete Werte für die Phasenlage erlaubt, strebt der Algorithmus zu oft einem lokalen Minimum zu und hat Probleme, ein globales Minimum zu finden.

Daher wurden im Laufe der Zeit eine Vielzahl von Untersuchungen und Verbesserungen an diesem Algorithmus durchgeführt. Besonders verdient gemacht hat sich J.R. FIENUP, der mit der Einführung eines Input-Output-Algorithmus-Kernel eine Möglichkeit geschaffen hat, auch für diskrete Phasen eine ausreichend gute Konvergenz zu erzielen⁴⁴.

Auch die Einführung einer gemächlichen Diskretisierung (engl. Soft Quantization) der Phase über mehrere Iterationsschritte durch Frank WYROWSKI löst das Problem der Konvergenz elegant auf. Daneben existieren noch viele andere Konzepte, wie z.B. eine Regularisierung, wie sie von KIM et. al. entwickelt wurde⁴⁵.

Eine vollständige Ausführung selbst dieser wichtigster Verbesserungen und Erweiterungen würde den Rahmen dieser Arbeit sprengen.

2.5 Technische Umsetzung der Design-Verfahren

2.5.1 Die MATLAB-Umgebung

Die präsentierten Verfahren und Prinzipien werden technisch mit Hilfe der Software MATLAB umgesetzt. Dies ist eine kommerziell von der Firma The MathWorks, Inc. entwickelte und vertriebene Rechen- und Analyseumgebung für überwiegend mathematische Problemstellungen.

MATLAB behandelt mathematische Größen und Operationen generell als Matrizen bzw. Matrizenoperationen. Daher leitet sich auch der Name von **matrix laboratory** ab. Skalare Größen werden als Sonderfall von Matrizen des Typs (1,1) angesehen. Diese Herangehensweise erlaubt eine schnelle Umsetzung und Berechnung auch mehrdimensionaler Probleme, ohne die zugrunde liegenden Algorithmen und Funktionen ändern zu müssen.

Eine eigene implementierte Programmiersprache ermöglicht die Nutzung eines umfangreichem Repertoires an vorhandenen Funktionen. Diese Sammlung reicht standartgemäß von einfachen Funktionen, z.B. zur Berechnung des Sinus, bis hin zu komplexen Algorithmen, z.B. zur Erzeugung von BESSEL-Funktionen oder Fast-FOURIER-Transformationen. Durch spezielle Toolboxes können noch weitere auf spezifische Problemstellungen ausgerichtete Funktionen und Elemente hinzugefügt werden (z.B. Bioinformatik Toolbox, Neuronale Netzwerke Toolbox, usw.).

⁴⁴ vgl. [Fienup]

⁴⁵ vgl. [Kim]

Berechnungen und Funktionen können zum einen in einem Kommandofenster ausgeführt bzw. aufgerufen werden. Zum anderen können sie auch in einem m-file (MATLAB-Datei) zusammengefasst werden. M-files ermöglichen so die Erzeugung komplexer zusammenhängender Algorithmen und Programme. Die Möglichkeit, Unterfunktionen zu erzeugen und einzubetten, ist dabei oft hilfreich. Auch das Programmmodul zur Berechnung von DOEs, das ein Ziel dieser Arbeit ist, wird als m-file ausgeführt.

MATLAB bietet zusätzlich zur Steuerung über das Kommandofenster die Variante grafischer Benutzeroberflächen (engl. graphical user interface, GUI). Diese erleichtern den Umgang mit komplexen Programmen erheblich, besonders die Parameterein- und -ausgabe. Im weiteren Verlauf dieser Arbeit wird auch eine GUI zur komfortablen Steuerung des ausgearbeiteten DOE-Berechnungsverfahrens erstellt.

2.5.2 Besonderheiten numerischer Berechnungen

Da die Umsetzung der beschriebenen Verfahren in MATLAB auf numerischem Weg erfolgt, sind allgemeine Numerik-Kenntnisse unumgänglich. Die Eigenheiten solcher Berechnungen sind in der programmtechnischen Realisierung zu beachten und geeignete Strategien sind anzuwenden, um grobe Fehler zu vermeiden. Im Folgenden werden einige dieser Fehler diskutiert.

Im Unterschied zu analytischen Rechentechniken, bei denen Ergebnisse auch mittels Variablen beschrieben werden können, beschränken sich numerische Verfahren auf Zahlen. Darüber hinaus können diese nur in einer endlichen Genauigkeit angegeben und behandelt werden. Genau darin liegt die Hauptursache für eine Vielzahl von Fehlern.

Gewöhnlich wird programmintern die Gleitkommadarstellung einer Zahl Z angewendet. Diese erlaubt es, sowohl sehr große als auch sehr kleine Zahlen mit gleicher relativer Genauigkeit darzustellen:

$$Z = \pm 0, z_1 z_2 z_3 \dots z_t \cdot 10^e$$

Bsp. $18,5 = 0,185 \cdot 10^1$

Für die Mantisse $z_1 z_2 z_3 \dots$ der normalisierten Gleitkommadarstellung gilt hierbei $z_1 \neq 0$ mit Ausnahme der Darstellung der Zahl 0^{46} . Die Länge t der Mantisse bestimmt gleichzeitig die Anzahl sicherer Ziffern einer Zahl. Die Verwendung des binären Zah-

⁴⁶ vgl. [Schwetlick], S. 17/ 18

lensystems in modernen Rechensystemen schränkt diese mögliche Genauigkeit zusätzlich ein. In dezimaler Schreibweise genügen für das Beispiel $t = 3$ sichere Ziffern, während im Binärsystem $t = 6$ nötig sind:

Bsp. $18,5_{\text{dez}} = 0,100101 \cdot 2^{101}$

Vor diesem Hintergrund ist festzustellen, dass Assoziativ- und Distributivgesetz nicht mehr uneingeschränkt gültig sind⁴⁷. Das folgende Beispiel demonstriert dies:

Bsp. $t = 8$

$$a = 0,23371258 \cdot 10^{-4} \quad b = 0,33678429 \cdot 10^2 \quad c = 0,33677811 \cdot 10^2$$

$$a + (b + c) = 0,23371258 \cdot 10^{-4} + 0,61800000 \cdot 10^{-3} = \underline{0,64137126 \cdot 10^{-3}}$$

$$(a + b) + c = 0,33678452 \cdot 10^2 - 0,33677811 \cdot 10^2 = \underline{0,64100000 \cdot 10^{-3}}$$

exakt: $a + b + c = 0,641371258 \cdot 10^{-3}$

Es ist also von Bedeutung, in welcher Reihenfolge bestimmte numerische Rechenoperationen durchgeführt werden. Unvermeidbare Subtraktionen sollten möglichst am Anfang der Rechnung stehen. Dies ist ein Punkt, der u.U. bei der Umsetzung beachtet werden muss.

Weiter können durch Stellenauslöschung bei Subtraktionsoperationen sichere Ziffern entfallen. Wie das nachfolgende Beispiel zeigt, kann dies die relativen Fehler erheblich erhöhen. Eine mögliche Gegenmaßnahme könnte das „Rechnen mit erhöhter Genauigkeit an kritischen Stellen“⁴⁸ sein.

Bsp.	<i>exakte Rechnung</i>	<i>mit $t = 4$ tragenden Ziffern</i>
	$42354 - 42356 = -2$	$0,4235 \cdot 10^5 - 0,4236 \cdot 10^5 = -10$

Wo dies möglich ist, sollten Ausdrücke äquivalent ersetzt werden, um Rechenfehler bei der Differenzenbildung zu vermeiden. Ein Beispiel dafür ist die Berechnung des Ausdrucks $1 - \cos x$ für $|x| \approx 2k\pi$. In der Nähe eines Vielfachen von 2π werden annähernd gleiche Zahlen subtrahiert, was durch die Rundungsfehler zu falschen Ergebnissen führt.

Bsp.	für $x = 6,28 (\approx 2\pi)$	$\cos 6,28 = 0,999994927$
------	-------------------------------	---------------------------

⁴⁷ vgl. [Stoer], S. 7

⁴⁸ [Bernert], S. 11

exakte Rechnung

$$1 - \cos 6,28 = 0,5073 \cdot 10^{-5}$$

mit $t = 4$ tragenden Ziffern

$$\cos 6,28 \approx 0,1000 \cdot 10^1$$

$$1 - \cos 6,28 = 0$$

Der Ersatz durch einen äquivalenten Ausdruck ohne Subtraktion ermöglicht eine weitaus genauere Berechnung:

$$1 - \cos x = \frac{1^2 - \cos^2 x}{1 + \cos x} = \frac{\sin^2 x + \cos^2 x - \cos^2 x}{1 + \cos x} = \frac{\sin^2 x}{1 + \cos x}$$

$$\sin 6,28 = 0,3185 \cdot 10^{-2}$$

$$(\sin 6,28)^2 = 0,1014 \cdot 10^{-4}$$

$$\frac{\sin^2 6,28}{1 + \cos 6,28} = \frac{0,1014 \cdot 10^{-4}}{2} = 0,507 \cdot 10^{-5}$$

Ein weiteres wichtiges Merkmal numerischer Algorithmen ist ihre Stabilität. Sowohl Rundungsfehler als auch Eingangsfehler sind in numerischen Verfahren unvermeidlich. Ein Algorithmus gilt dann als numerisch stabil, wenn die Rundungsfehler stets um einen Faktor C kleiner sind als die Eingangsfehler. Das Verfahren ist umso stabiler, je kleiner C ist. Andernfalls können sich Rundungsfehler in einem Algorithmus aufsummieren und zu falschen Ergebnissen führen.

Im folgenden Beispiel soll zur Verdeutlichung dieser Aussage eine Zahlenfolge errechnet werden:

Bsp. $I_n = \frac{1}{e} \cdot \int_0^1 x^n e^x dx$ für $n = 0, 1, 2, \dots, N$

Es ist möglich, einen Startwert mit $I_0 = 1 - \frac{1}{e}$ zu bestimmen und mittels partieller Integration folgende Rekursionsformel zur Berechnung der Reihenglieder herzuleiten:

$$I_n = 1 - n \cdot I_{n-1} \quad \text{für } n = 1, 2, \dots, N$$

Es ergeben sich folgende Werte:

n	I_n (exakt)	I_n^* (numerisch, 2. Stelle nach Komma)
0	0,63	0,63
1	0,37	0,37
2	0,26	0,26
3	0,21	0,22
4	0,17	0,12

5	0,15	0,40
6	0,13	-1,40
7	0,11	10,80
8	0,10	-85,40

Bis zu einem gewissen Punkt sind die Ergebnisse noch brauchbar, doch dann überwiegen die numerischen Rundungsfehler. Selbst ein deutlich genauer angegebener Anfangswert würde dies nicht ändern. Der kritische Punkt verschiebt sich nur weiter nach hinten⁴⁹. Ein solcher Algorithmus ist numerisch instabil und damit nicht nutzbar.

2.5.3 Die Fast FOURIER Transformation (FFT)

Besondere Aufmerksamkeit für eine rechentechnische Umsetzung der beschriebenen Methoden muss der numerischen Berechnung der FOURIER-Transformation gewidmet werden. Setzt man das bekannte FOURIER-Integral

$$F(k) = \int_{-\infty}^{\infty} f(x) \cdot e^{-ikx} dx \quad (2-53)$$

einfach nur in eine numerische Form um, würde sich diese Summe wie folgt ergeben:

$$F(k_n) = \sum_{m=0}^{M-1} [f(x_m) \cdot e^{-i \cdot k_n \cdot x_m} \cdot (x_{m+1} - x_m)] \quad ; n = 0, 1, \dots, N-1 \quad (2-54)$$

Das zu transformierende Signal $f(x_m)$ wird an M Stellen x_m abgetastet und zur Berechnung eines Frequenzwertes k_n herangezogen. Die dabei durchzuführende komplexe Multiplikation benötigt den größten Rechenaufwand. Da diese Operation für N k -Werte durchgeführt werden muss, ist die Berechnungsdauer direkt proportional zu $M \cdot N$. Von mathematischen Gesichtspunkten her ist es sinnvoll, M und N möglichst gleich zu wählen. Andernfalls wird die Funktion $f(x)$ zwar eng abgetastet und die Frequenzwerte k_n werden recht genau berechnet, aber es werden trotzdem kaum mehr Informationen im Frequenzraum gewonnen. Der entgegengesetzte Fall wäre, dass $N \gg M$ ist. Dies führt dazu, dass zwar viele Frequenzen k_n ermittelt werden, sie jedoch nur sehr ungenau berechnet werden.

⁴⁹ vgl. [Bernert], S. 16

Das breite Anwendungsfeld der FOURIER-Transformation über die FOURIER-Optik hinaus machte daher schon bald nach dem Aufkommen moderner Rechentechnik einen Algorithmus nötig, der eine deutliche Reduzierung der Rechendauer ermöglicht. 1965 veröffentlichten COOLEY und TUKEY einen Algorithmus, der als „Fast FOURIER Transformation“ bekannt wurde. Voraussetzung für dessen Einsatz ist eine Zweierpotenz-Diskretisierung der Ausgangsfunktion, d.h. $N = 2^\gamma$ ($\gamma \in \mathbb{N}$). Diese Methode ermöglicht eine Berechnungsdauer, die direkt proportional zu $N \cdot \log_2 N$ ist.⁵⁰

Eine komplette, mathematisch korrekte Herleitung und Erläuterung des FFT-Algorithmus würde den Rahmen dieser theoretischen Vorbetrachtung sprengen. Es sei an dieser Stelle auf [Brigham]⁵¹ verwiesen, wo diese Methode detailliert beschrieben wird. Stattdessen soll nur kurz am Beispiel $N = 2^2 = 4$ das Grundprinzip verständlich gemacht und die Rechenzeiteinsparung erklärt werden.

Ausgangspunkt der FFT ist eine verkürzte Schreibweise der Diskreten FOURIER-Transformation (DFT):

$$F(n) = \sum_{m=0}^{N-1} [f_o(m) \cdot e^{-i \cdot 2\pi \cdot n \cdot m / N}] \quad ; n = 0, 1, \dots, N-1 \quad (2-55)$$

Diese Schreibweise ähnelt zwar der numerischen Formulierung des kontinuierlichen Integrals (2-54), unterscheidet sich von ihr aber auch in entscheidenden Punkten. Einerseits laufen sowohl m als auch n von 0 bis N , das heißt, Raumbereich und Frequenzbereich sind fest miteinander verknüpft.

Andererseits ist in (2-54) nirgendwo festgelegt, an welchen Stellen x_m die Funktion abgetastet werden muss bzw. welche Frequenzen k_n berechnet werden müssen. Diese Stützstellen sind prinzipiell frei wählbar. In (2-55) dagegen erfolgt die Abtastung im Frequenz- und Raumbereich von 0 ausgehend an äquidistanten Stützstellen.

Das Problem des Informationsverlustes, der offenbar durch die Vernachlässigung der negativen Anteile entsteht, lässt sich anhand der mathematischen Herleitung der DFT lösen. Da für die DFT nur ein endliches Intervall der Funktion $f(x)$ betrachtet werden kann, „nimmt die DFT an“, dass es sich um eine periodische Funktion handelt (siehe Abbildung 18). Die Verschiebung des Abtastintervalls in den positiven Bereich führt dementsprechend dazu, dass die scheinbar fehlenden negativen Funktionsteile als nächste Periode einbezogen werden. Analoges kann für den Frequenzbereich hergeleitet werden⁵².

⁵⁰ vgl. [Brigham] S. 21

⁵¹ [Brigham], S. 181-219

⁵² vgl. [Brigham], S. 116-122

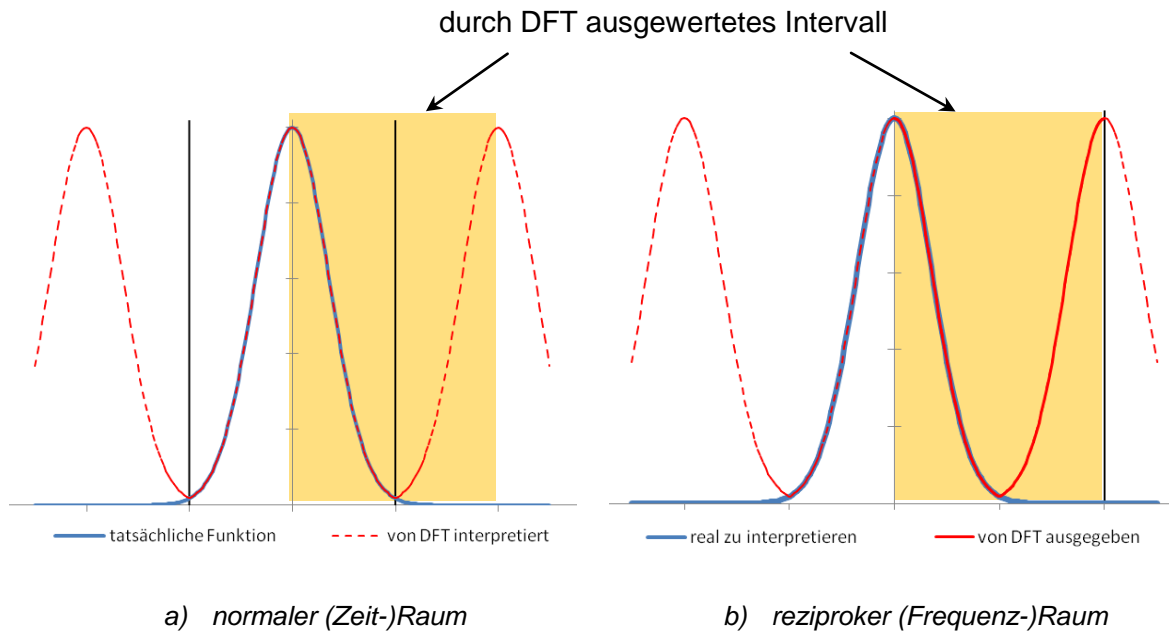


Abbildung 18 Periodizität von Funktionen bei der DFT

Die so entstehende scheinbare Verschiebung in der FOURIER-Transformierten wird auch als „Frequenz-Shift“ bezeichnet.

Für $N = 4$ Diskretisierungswerte ergibt sich aus (2-55) ein Gleichungssystem der folgenden Form:

$$\begin{aligned}
 F(0) &= f_o(0) \cdot W^0 + f_o(1) \cdot W^0 + f_o(2) \cdot W^0 + f_o(3) \cdot W^0 & (2-56) \\
 F(1) &= f_o(0) \cdot W^0 + f_o(1) \cdot W^1 + f_o(2) \cdot W^2 + f_o(3) \cdot W^3 \\
 F(2) &= f_o(0) \cdot W^0 + f_o(1) \cdot W^2 + f_o(2) \cdot W^4 + f_o(3) \cdot W^6 \\
 F(3) &= f_o(0) \cdot W^0 + f_o(1) \cdot W^3 + f_o(2) \cdot W^6 + f_o(3) \cdot W^9
 \end{aligned}$$

Der Faktor W steht darin jeweils vereinfachend für die konstante e-Funktion:

$$W = e^{-i \cdot 2\pi/N} \quad (2-57)$$

Dieses Gleichungssystem lässt sich in Matrizenschreibweise zusammenfassen zu:

$$\begin{pmatrix} F(0) \\ F(1) \\ F(2) \\ F(3) \end{pmatrix} = \begin{bmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{bmatrix} \cdot \begin{pmatrix} f_o(0) \\ f_o(1) \\ f_o(2) \\ f_o(3) \end{pmatrix} \quad (2-58)$$

Für die Komponenten W gilt der Zusammenhang

$$W^{n \cdot m} = W^{n \cdot m \bmod(N)}, \quad (2-59)$$

so dass sich (2-58) vereinfachen lässt zu:

$$\begin{pmatrix} F(0) \\ F(1) \\ F(2) \\ F(3) \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W^1 & W^2 & W^3 \\ 1 & W^2 & W^0 & W^2 \\ 1 & W^3 & W^2 & W^1 \end{bmatrix} \cdot \begin{pmatrix} f_o(0) \\ f_o(1) \\ f_o(2) \\ f_o(3) \end{pmatrix} \quad (2-60)$$

Anschließend erfolgt eine Zerlegung der Koeffizientenmatrix in ein Matrizenprodukt wie nachfolgend dargestellt. Der Exponent γ zeigt die Anzahl der nötigen Einzelmatrizen. Zur genauen Herleitung dieser Faktorisierung sei erneut auf [Brigham]⁵³ verwiesen.

$$\begin{pmatrix} F(0) \\ F(2) \\ F(1) \\ F(3) \end{pmatrix} = \begin{bmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{bmatrix} \cdot \begin{pmatrix} f_o(0) \\ f_o(1) \\ f_o(2) \\ f_o(3) \end{pmatrix} \quad (2-61)$$

Es wird hier darauf hingewiesen, dass durch die Zerlegung der Matrix in der oben beschriebenen Weise die Reihenfolge der Zeilen im Ergebnisvektor \vec{F} vertauscht werden müssen. Dieser Wechsel erfolgt im Allgemeinen anhand der gespiegelten binären Indizes (z.B. wird aus $10_{\text{bin}} = 2_{\text{dez}}$ demnach $01_{\text{bin}} = 1_{\text{dez}}$).

Diese Schreibweise der Indizes als Binärzahlen zum Tauschen der Zeilen ist eine der Gründe für die Zweierpotenz der Abtastwerte $N = 2^\gamma$. Weicht N von dieser Bedingung ab, ergeben die Spiegelungen der binären Indizes Zeilennummern, die nicht mehr zugeordnet werden können.

Wie bereits mit (2-59) gezeigt, resultiert die Rechenzeitverkürzung hauptsächlich aus der Einsparung von komplexen Multiplikationen, indem gesuchte Ergebnisse auf bereits berechnete zurückgeführt werden. Die anschließenden Zwischenberechnungen der Matrizen verdeutlichen dies nochmal.

$$\begin{pmatrix} f_1(0) \\ f_1(1) \\ f_1(2) \\ f_1(3) \end{pmatrix} = \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{bmatrix} \cdot \begin{pmatrix} f_o(0) \\ f_o(1) \\ f_o(2) \\ f_o(3) \end{pmatrix} \quad (2-62)$$

Es kann gezeigt werden, dass gilt:

$$W^2 = -W^0 \quad (2-63)$$

Damit werden zur Berechnung des Zwischenergebnisses \vec{f}_1 nur zwei statt vier komplexe Multiplikationen benötigt. In ähnlicher Weise kann man die Multiplikation der verbliebenen Matrix auswerten.

⁵³ [Brigham], S. 209-212

$$\begin{pmatrix} F(0) \\ F(2) \\ F(1) \\ F(3) \end{pmatrix} = \begin{pmatrix} f_2(0) \\ f_2(1) \\ f_2(2) \\ f_2(3) \end{pmatrix} = \begin{bmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{bmatrix} \begin{pmatrix} f_1(0) \\ f_1(1) \\ f_1(2) \\ f_1(3) \end{pmatrix} \quad (2-64)$$

Es zeigt sich, dass auch diese Operation mit zwei komplexen Multiplikationen und vier komplexen Additionen auskommt. Insgesamt erfordert die Berechnung von \vec{F} nach dieser Methode vier komplexe Multiplikationen statt der $4^2 = 16$ Multiplikationen, die eine Berechnung nach (2-56) erfordert. Anschließend ist noch ein Umsortieren der Zeilen in \vec{F} notwendig.

Eine allgemeine Darstellung des Algorithmus zeigt, dass für die Berechnung der FOURIER-Transformation mittels FFT $N \cdot \gamma / 2$ komplexe Multiplikationen nötig sind, während die direkte Methode N^2 Multiplikationen benötigt. Es kann insbesondere für größere N davon ausgegangen werden, dass die Gesamtrechenzeit maßgeblich von den Multiplikationen bestimmt wird. Der Rechenzeitvorteil der FFT lässt sich damit näherungsweise angeben mit

$$\frac{N \cdot \gamma / 2}{N^2} = \frac{\gamma}{2N} \quad (2-65)$$

Ein Struktogramm des kompletten allgemeinen Algorithmus nach [Brigham] ist in Anlage 1 zu finden.

3 Umsetzung des Programmmoduls

3.1 Vorbetrachtungen

Bevor der Berechnungsalgorithmus in MATLAB umgesetzt werden kann, muss über verschiedene grundlegende Aspekte Klarheit bestehen. Es wird entschieden, welcher der vorgestellten Algorithmen implementiert wird, welche Daten einzugeben sein müssen und können und in welcher Form die Ergebnisse dem Nutzer wieder zugänglich gemacht werden.

3.1.1 Zur Wahl des umzusetzenden Verfahrens

In den folgenden Abschnitten wird die Umsetzung des iterativen FOURIER-Transformations-Algorithmus (IFTA) in ein MATLAB-Programmmodul zur Berechnung von DOEs dargestellt.

Bei den Recherchen nach Möglichkeiten zur DOE-Berechnung ist festzustellen, dass es kaum ein Verfahren in diesem Bereich gibt, über das so viel geforscht und veröffentlicht wurde. Allein eine Sammlung aller Abwandlungen, Optimierungen und Varianten des IFTA zusammenzutragen und auszuwerten, wäre Stoff für eine eigenständige Arbeit. Es handelt sich also nicht um einen abgeschlossenen Algorithmus, sondern die ständigen Weiterentwicklungen halten ihn trotz seiner fast 40jährigen Geschichte aktuell.

Dies, die verhältnismäßig einfache Implementierung und die Nutzung aktueller IFTA-Abwandlungen in den aktuellen Optikdesign-Programmen waren ausschlaggebend für die Entscheidung für den IFTA. Eine eigene rechentechnische Umsetzung des Algorithmus wird dem Verständnis der Prozesse innerhalb solcher Software förderlich sein, was die Nachvollziehbarkeit von Rechenergebnissen ermöglicht.

3.1.2 Festlegung von Eingangsparametern

Die ersten Überlegungen zur praktischen Umsetzung des IFTA beziehen sich auf die Eingangsparameter. Es muss geklärt werden, welche Parameter wie und in welcher Form vom Nutzer abgefragt werden.

Dem Eingangsfeld kommt besondere Bedeutung zu. Wie auch in späteren Untersuchungen gezeigt wird (siehe 4.3.2), muss dieses exakt festgelegt sein, da das ge-

plante DOE darauf ausgerichtet sein wird. Prinzipiell ist hier jede beliebige Verteilung möglich, jedoch soll das Programmmodul auf eine gaußverteilte Eingangintensität ausgelegt sein. Mittels der Variablen I_0 für die Spitzenintensität und w für den Strahlradius des Gaußstrahls kann dieser den Anforderungen des Nutzers angepasst werden.

Auch die Form der Wellenfronten kann flexibel ausgelegt sein. Es ist im Algorithmus möglich, z.B. auch sphärische Wellenfronten einer Punktquelle zu berücksichtigen. Auch diesbezüglich soll die Umsetzung einfach gehalten bleiben und sich auf ebene Wellenfronten mit der Phasenlage 0 beschränken. Die häufige Verwendung der DOEs in Verbindung mit kollimierten Laserstrahlen macht diese Vereinfachung auch praktisch sinnvoll.

Die Modellierung nicht- oder teilkohärenter Strahlungsquellen (Bsp. Excimerlaser) ist zwar möglich, jedoch eine große Herausforderung⁵⁴. Dies kann in späteren Erweiterungen des Moduls eingeplant werden. In dieser Umsetzung soll nur kohärente Strahlung berücksichtigt werden.

Die Wellenlänge λ ist für den optischen Aufbau des DOEs äußerst wichtig. Schließlich werden diese i.d.R. für eine bestimmte Wellenlänge designt, für die diese dann die besten Resultate liefern. Für den ausgearbeiteten Algorithmus spielt sie jedoch keine Rolle. Da im IFTA mit komplexen Amplituden gerechnet wird, können die ermittelten DOEs prinzipiell für alle Wellenlängen genutzt werden. Erst, wenn die berechnete Phasenmodulierung in ein konkretes Höhenprofil umgewandelt werden muss, kommt die Wellenlänge und mit ihr alle anderen verbundenen Größen (Brechzahl etc.) ins Spiel. Alle Größen, die vorher im Algorithmus numerisch verwendet werden, bekommen ihre physikalische Entsprechung erst mit Festlegung einer Arbeitswellenlänge. Für die Realisierung in MATLAB ist diese Größe also uninteressant.

Wichtiger ist eine Angabe über die Genauigkeit der Berechnung. Es sind in diesem Zusammenhang Werte für die Diskretisierungen N und p vom Nutzer abzufragen. N sollte dabei nach Möglichkeit eine Potenz von 2 bilden, da dann der FFT-Algorithmus die größte Effizienz aufweist und so am schnellsten arbeitet. Die Diskretisierung der Phasenwerte p legt die Anzahl an Werten zwischen $-\pi$ und π fest, die vom Algorithmus genutzt werden dürfen.

Die größte Bedeutung unter den Eingabedaten kommt natürlich der gewünschten Intensitätsverteilung zu. Sie ist für den Verlauf des Prozesses entscheidend. Unter Berücksichtigung der spezifischen Strahlprofil-Anforderungen in der Lasertechnik sollten folgende Zielverteilungen zur Auswahl stehen:

⁵⁴ siehe [Wyrow2007]

- Top Hat – Profil
- Linien – Profil
- Strahlteiler

Es sind für den Benutzer zu jeder dieser Auswahlmöglichkeiten auch die passenden Eingabemöglichkeiten vorzusehen, z.B. Linienbreite und Ausrichtung für das Linienprofil. Um eine hohe Flexibilität hinsichtlich berechenbarer Strahlverteilungen zu gewährleisten, sollte zusätzlich eine Funktion zum Einlesen unregelmäßiger Profile existieren. Mögliche Dateiformate sind ASCII-Textdateien oder Bitmap-Bilddateien.

Letztlich sollte noch die Auswahlmöglichkeit zwischen den Iterationsmodi „Schritt“ und „Speed“ gegeben sein. Der Schrittmodus erfordert für jeden weiteren Iterationsschritt eine Aktion, wie z.B. der Bestätigung über einen Button. Dies ist besonders für Demonstrations- und Kontrollzwecke nützlich, bei denen die Entwicklung des Simulationsergebnisses über die Iterationsschritte verfolgt werden muss.

Die Speed-Funktion führt demzufolge eine eingegebene Zahl von Iterationsschritten aus und stellt nur noch das Endergebnis dar.

3.1.3 Ausgabe der Ergebnisse

Geht es um die praktische Nutzung der gewonnenen Erkenntnisse, stellt sich die Frage nach den Ausgabeformaten der Daten.

Zur Validierung des Verfahrens-Ablaufes ist eine Möglichkeit der ständigen Ergebniskontrolle vorzusehen. Die einfachste Variante ist es, während der einzelnen Iterationsschritte die jeweils berechnete Ausgangsintensität darzustellen. So kann der Fortschritt der Rechnung hin zur gewünschten Zielverteilung während des Prozesses kontrolliert werden. Da in modernen Computern die Abarbeitung dieses Algorithmus sehr schnell abläuft, ist unter Umständen eine Möglichkeit zur Schrittsteuerung vorzusehen.

Neben der erreichten Intensitätsverteilung, welche in einer geeigneten grafischen Darstellung angezeigt werden sollte, ist die Ausgabe der errechneten Daten nötig. Besonders im Hinblick auf die Verwendung der errechneten Phasenverteilungen zur praktischen Herstellung der DOE oder zur Nachsimulation des Strahlverlaufes ist diese Ausgabe notwendig. Denkbare Formate, die von vielen Programmen zum Import unterstützt werden, sind ASCII- und Bitmap-Dateien. Beide Formate ermöglichen die exakte Zuordnung berechneter Phasenwerte zu den jeweiligen Pixeln. Beim Bitmap-Format ist zu beachten, dass in Graustufen nur 256 verschiedene Werte zur Verfügung stehen.

3.2 Entwicklung des Programmes

Die wichtigsten Schritte der Struktogramm-Entwicklung und GUI-Erstellung werden erläutert. Während des Programmierens aufgetretene Probleme werden mit ihrer Lösung dargestellt.

3.2.1 Grafische Benutzeroberfläche

Die grafische Benutzeroberfläche (auch: GUI, von **G**raphical **U**ser **I**nterface) bietet einen schnellen und bedienungsfreundlichen Zugang zu allen wichtigen Funktionen und Parametern des programmierten Moduls. Abbildung 19 zeigt die Form, die unter Berücksichtigung der Überlegungen aus 3.1.2 und 3.1.3 entstanden ist.

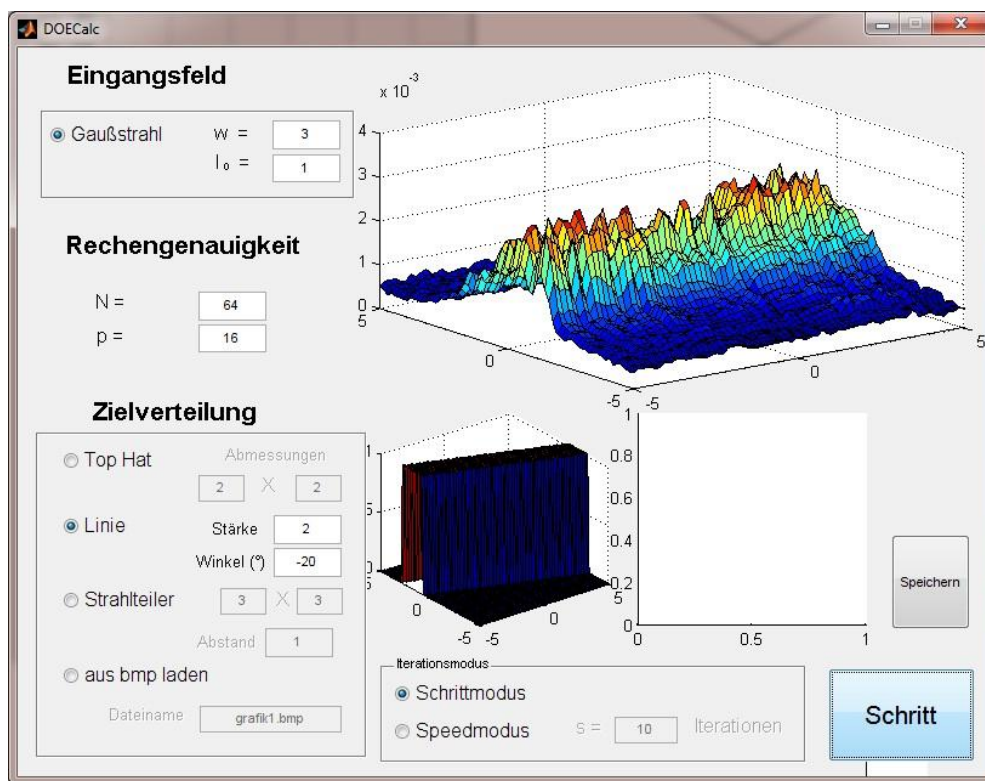


Abbildung 19 GUI des entwickelten MATLAB-Moduls

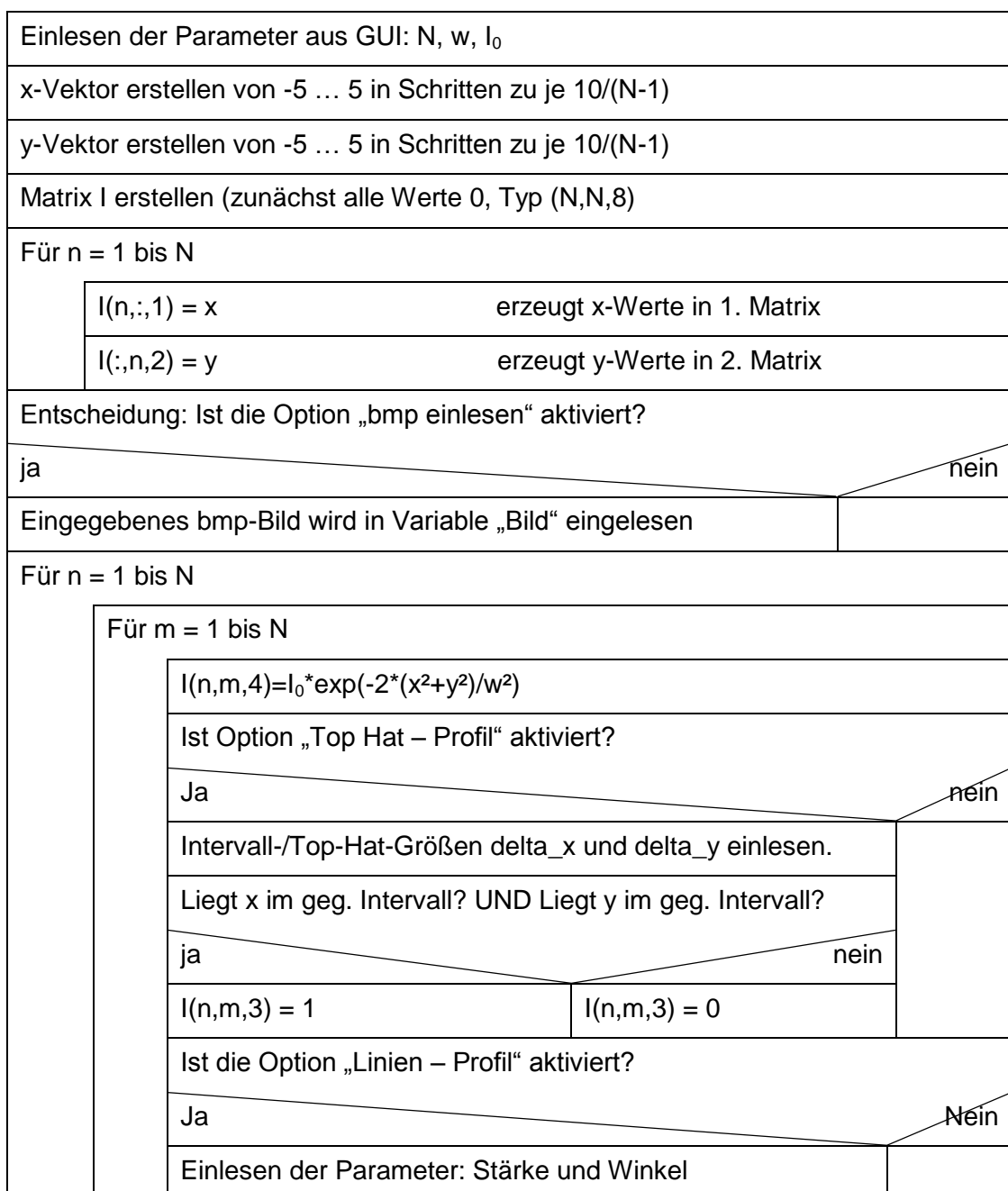
Die Auswahlmöglichkeit im Bereich „Eingangsfeld“ wird schon in Vorbereitung auf eventuell folgende Erweiterungen angelegt, in denen der Nutzer die Wahl zwischen verschiedenen Strahlquellen bekommt.

Es wird angestrebt, dass sich die Oberfläche dem Nutzer anpasst. So wird z.B. die Aufschrift des Startbuttons, der den Algorithmus in Gang setzt, bei aktiviertem Schrittmodus in „Schritt“ geändert. Zudem sind für eine bestimmte Auswahl nicht relevante Eingabefelder grau und inaktiv und werden erst mit Auswahl des entsprechenden Feldes aktiviert.

3.2.2 Struktogramm

Das Programmmodul besteht im Wesentlichen aus zwei Hauptbestandteilen, deren Struktogramme im Folgenden dargestellt und erklärt werden sollen.

Das erste Struktogramm in Abbildung 20 zeigt die Funktion „Set_Matrix“. Diese dient dazu, Änderungen, die der Nutzer an den Parametern vornehmen kann, sofort in den Datensatz aufzunehmen und diesen dementsprechend zu aktualisieren. Zur Funktionsweise der großen Datenmatrix I wird auf 3.2.3 verwiesen. Im Endeffekt wird das Wirken dieser Funktion dadurch deutlich, dass fast unmittelbar nach Änderung eines Parameters der Eingangsdaten die neu angepasste Zielverteilung angezeigt wird. Je nach numerischem Aufwand kann es einige Zeit dauern, bis alle Punkte berechnet worden sind.



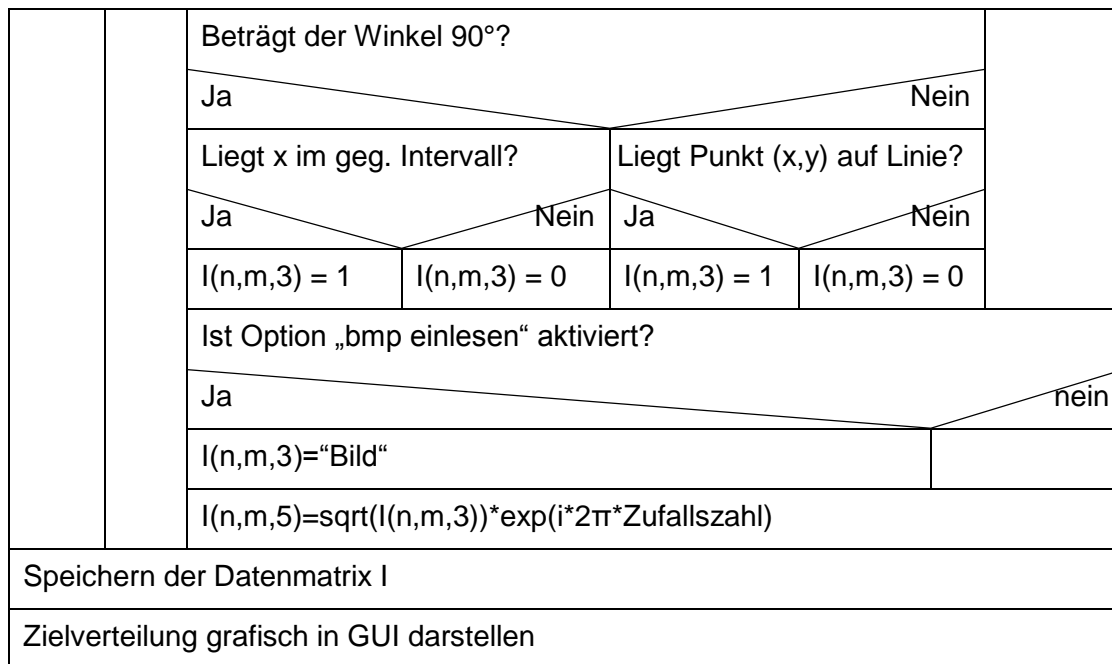
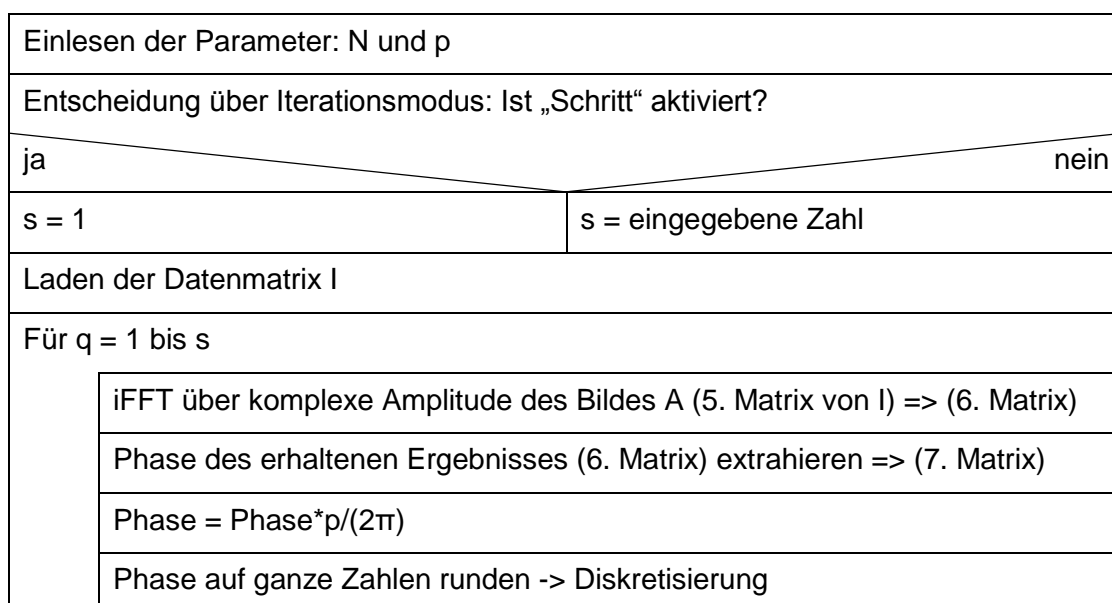


Abbildung 20 Struktogramm der Funktion "Set_Matrix"

In der FOR-Schleife wird jedem Punkt (x,y) gemäß der ausgewählten Zielverteilung sein Intensitätswert zugeordnet. Dass innerhalb der Schleife nur zwischen den drei Fällen Top-Hat, Linie und Bild unterschieden wird, bedeutet, dass die Erzeugung eines Strahlteiler-Array aktuell nur über Einlesen einer bmp funktioniert. Darüber hinaus wird bereits hier für jeden Punkt die komplexe Startamplitude mit einer zufälligen Phase berechnet.

Der wichtigere zweite Teil des Programms besteht im eigentlichen iterativen Algorithmus. Dieser kann sehr kompakt gehalten werden, da in MATLAB bereits fertige Funktionen zur Durchführung der zweidimensionalen FOURIER-Transformation sowie ihrer Inversen zur Verfügung stehen. Das in Abbildung 21 abgebildete Struktogramm kann größtenteils direkt aus dem Prinzip des IFTA abgeleitet werden.



	Phase = Phase $\cdot (2\pi)/p$
	Amplitude in 7. Matrix mit geg. Eingangsamplitude ersetzen
	FFT über komplexe Amplitude des DOEs durchführen => (8. Matrix)
	Betragsquadrat der Elemente der 8. Matrix grafisch darstellen
	Betrag der Amplitude der 8. Matrix mit Zielverteilung ersetzen
	Speichern der neuen Datenmatrix I

Abbildung 21 Struktogramm der Funktion IFTA

Auch dieser zweite Modulbestandteil bedient sich der Datenmatrix I.

Es gibt noch einige kleinere Funktionen, die fast ausschließlich die Aufgabe haben, eine der beiden vorgestellten Hauptfunktionen aufzurufen, wenn ein bestimmter Button angeklickt wurde oder der Wert eines Textfeldes geändert wurde. Es wird darauf verzichtet, diese alle aufzuführen.

Der MATLAB-Code beider Programmteile ist in Anhang 2 bereitgestellt.

3.2.3 Anmerkungen zur Programmierung

Im Programmmodul wurden die Daten in besonderer Form behandelt. Beim Wechsel von einer Funktion in die andere können Daten normalerweise übergeben werden. Da dies allerdings erfordert, in beiden Funktionen exakt die Form und Menge der Daten festzulegen, die übermittelt werden, schleichen sich sehr schnell Fehler in der Programmierung ein. Im vorliegenden Modell wurde daher der Weg gewählt, sämtliche Daten in einer externen Datei im Arbeitsverzeichnis zu speichern und bei Bedarf wieder zu laden. So können unterschiedliche Funktionen die Daten verändern, ohne sich dabei gegenseitig zu stören.

Diese Methode ist zwar recht übersichtlich, hat allerdings den Nachteil, auf die Schreib- und Lesezeiten der Festplatte angewiesen zu sein. Gerade bei großen Datenmengen kann dies ein limitierender Faktor werden.

Die Daten, mit denen im IFTA gerechnet wird, sind alle in einer großen dreidimensionalen $N \times N \times 8$ -Datenmatrix I zusammengefasst. Diese Matrix ist so beschaffen, dass ihre einzelnen Ebenen jeweils andere Informationen beinhalten:

1. Matrix: x-Werte
2. Matrix: y-Werte
3. Matrix: Verteilung der Zielintensität
4. Matrix: Verteilung der Eingangsintensität
5. Matrix: komplexe Amplitude A des Beugungsbildes (korrigiert)

- 6. Matrix: komplexe Amplitude a des Bildes auf dem DOE
- 7. Matrix: komplexe Amplitude a' auf dem DOE nach Korrektur
- 8. Matrix: komplexe Amplitude A' des Beugungsbildes

Auf diese Weise sind die Informationen eines jeden diskretisierten Punktes (n,m) in der Matrix abrufbar, jede in einer anderen Ebene. Die Funktion des eigentlichen IFTA ändert in ihrem Verlauf nur die komplexen Amplituden, d.h. die Matrizen 5-8. Eine Änderung der Eingangsdaten wirkt sich nur auf die ersten vier Matrizen aus.

Weiter ist bei der Ausgabe der Phase in eine Bitmap-Bilddatei das spezielle Format zu beachten. Die MATLAB-Funktion `imwrite` erzeugt automatisch aus den Daten einer Matrix eine solche Bitmap-Datei, indem sie den Daten zwischen 0 und 1 jeweils einen Grauwert von 0 bis 255 zuordnet. Da die Phase zum einen auch negativ und zum anderen bis zu einem Wert von π angegeben ist, muss sie zunächst in den positiven Bereich verschoben und anschließend auf 2π normiert werden. Erst dann ergibt sich das korrekte Bild.

3.2.4 Verbesserungspotential

Da der Zweck des präsentierten Moduls die grundsätzliche Realisierung von DOE-Berechnungen ist, besteht noch jede Menge Raum für Verbesserungen. Diese beginnen bei der Einarbeitung von Eingabesicherungen, die Falscheingaben vermeiden, und endet bei der Implementierung weiterer Algorithmen. Diese können den IFTA ergänzen und Effizienz und Geschwindigkeit erhöhen oder sie können eine alternative Berechnungsmethode bieten.

Konkrete Verbesserungsvorschläge sind die Einführung von Kriterien für die Güte eines DOEs und darauf basierende automatisierte Abbruchkriterien. Dies würde die Handhabbarkeit des Verfahrens für weniger erfahrene Nutzer verbessern. Ebenso erhöht z.B. eine Vorschau beim Import von bmp-Dateien die Bedienfreundlichkeit.

Verbesserungspotential besteht auch bei den verfügbaren Strahlquellen. Hier könnten reale nicht- oder teilkohärente Quellen eingebaut werden. In diesem Zuge sollte ein realer Bezug zu absoluten Größen, wie z.B. der Wellenlänge, dem Strahldurchmesser und der Pixelgröße hergestellt werden. Die ist in der vorgestellten Version nicht nötig, würde aber das Verständnis der physikalischen Zusammenhänge fördern. Nicht zu vergessen ist die Berechnung eines konkreten Höhenprofils, das als Grundlage für die praktische Herstellung dienen kann.

Neben den vielen kleinen ungelösten Problemen, sollte zukünftig die Quantisierung der berechneten Phasenverteilungen realisiert werden. Aktuell wird diese Quantisierung allein durch die numerische Grenze des Algorithmus erzielt. Die Pixel ergeben sich aus der begrenzten räumlichen Auflösungsfähigkeit der Berechnung.

Hier ist eine Möglichkeit zu schaffen, die Pixelgröße z.B. an Herstellungsbedingungen orientieren zu können.

Zusätzlich zur grafischen Ausgabe des Rechenergebnisses können noch weiter Auswertungs-Werkzeuge bereitgestellt werden. Diese können numerische Werte liefern, wie z.B. die Konversionseffizienz oder die in einem Profil eingeschlossene Leistung. Sie könnten aber auch Querschnitte der erzeugten Verteilung generieren.

4 Theoretische Verifizierung der Ergebnisse

4.1 Testergebnisse

Es werden simulierte Strahlprofile vorgestellt und ausgewertet. Die gestellten Aufgaben wurden z.T. der Lasermaterialbearbeitung entnommen. Es wird gezeigt, dass der programmierte Algorithmus offenbar nicht alle geforderten Strahlformen in gleich guter Art realisieren kann.

4.1.1 Strahlformung: Top-Hat-Profil

Die erste zu realisierende Aufgabe ist die Transformation eines Gaußstrahls in ein Top-Hat-Profil. Dies ist eine in der Lasertechnik überaus häufig anzutreffende Aufgabenstellung, da für bestimmte Laseranwendungen, wie z.B. Laserauftragschweißen oder Laserhärten, eine solche räumliche Leistungsverteilung angestrebt wird. Eine DOE-basierende Strahlformung erspart kostspielige und platzraubende Optiksyste me aus mehreren Optiken.

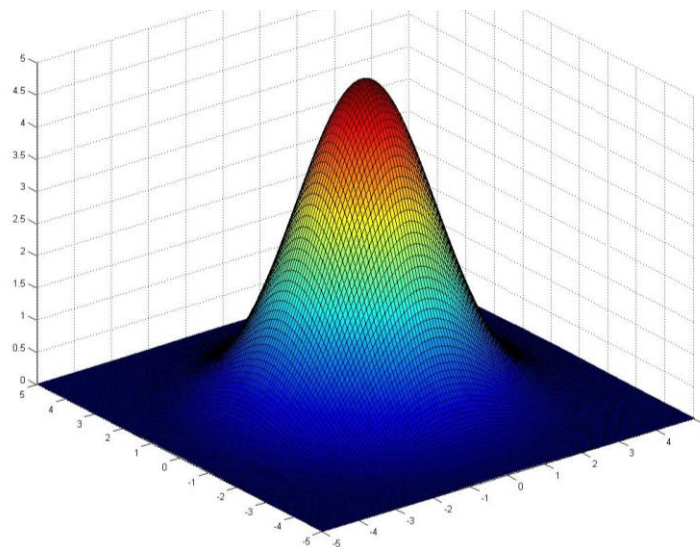


Abbildung 22 Gaußverteilung als simulierter Eingangsstrahl
(erstellt mit MATLAB)

Abbildung 22 zeigt das modellierte Eingangsprofil. Diese räumliche Gauß-Verteilung wird auch in allen folgenden Simulationen als Eingangsstrahl definiert. Es werden ebene Wellenfronten angenommen, so dass die einfallenden Wellen eine Phase von 0 haben.

Wie auch in den meisten nachfolgenden Untersuchungen wurden die berechneten Phasen in $p = 16$ Stufen gequantelt und jede Raumrichtung mit $N = 2^7 = 128$ Punk-

ten diskretisiert. 10 Iterationen des Algorithmus waren in der Regel ausreichend, um ein ansprechendes Ergebnis zu erhalten. In Abbildung 23 ist dargestellt, wie die simulierte Ausgangsintensität mit den jeweiligen Iterationsschritten der gewünschten Zielverteilung angenähert wird.

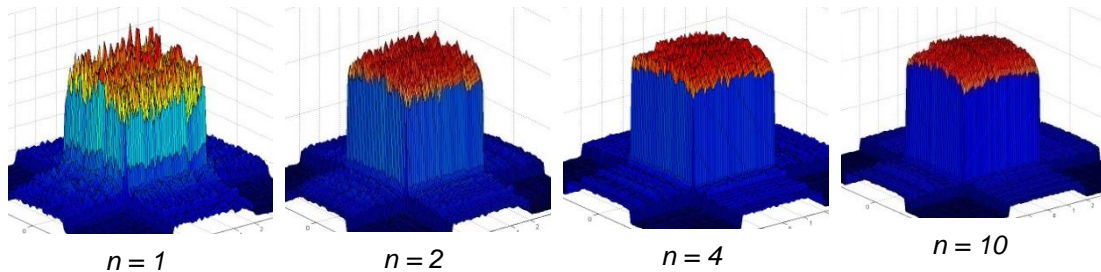


Abbildung 23 Simulierte Ausgangsverteilung konvergiert zum gewünschten Strahlprofil (erstellt mit MATLAB)

Es ist erkennbar, dass die geforderte Intensitätsverteilung in guter Näherung erreicht werden konnte. Auch die sehr scharfen Kanten des Profils und der damit verbundene hohe Kontrast zeugen von einer erfolgreichen Simulation. Anhand dieser Ergebnisse konnte auch die Eignung des Algorithmus für die Berechnung solcher Aufgaben unter Beweis gestellt werden. In späteren Beispielen wird gezeigt, dass dies nicht immer zutrifft.

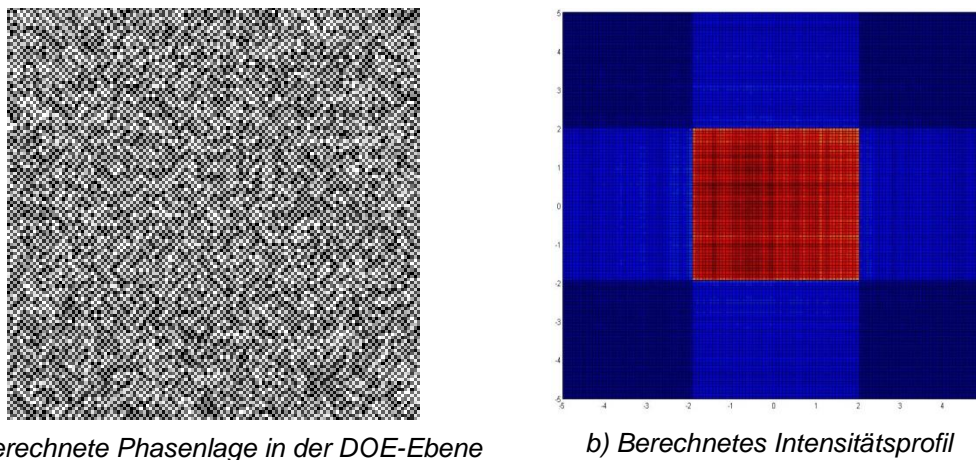


Abbildung 24 Simuliertes Ergebnis: Top-Hat-Profil (erstellt mit MATLAB)

Abbildung 24b zeigt das Ergebnis nach 10 Iterationen. Es sind seitlich des erzeugten Top-Hat-Profiles sowohl in x- als auch in y-Richtung erhöhte Beugungseffekte zu erkennen, die zu einer geringfügig erhöhten Intensität in diesen Bereichen führen. Verglichen mit der Intensität im Top-Hat ist dies jedoch marginal. Die Ursachen für diese Erscheinungen sollten in weiteren Untersuchungen geklärt und, wenn möglich, reduziert werden. Denkbar sind z.B. Überlagerungen höherer Beugungsordnungen. Auf diesem Weg könnte die Verlustleistung, die für den Top-Hat nicht zur Verfügung steht, minimiert und die Effizienz der Profil-Umwandlung erhöht werden.

4.1.2 Strahlformung: Linien in x- und y-Richtung

Linienprofile sind ebenfalls sehr gebräuchlich. Neben verschiedenen Materialbearbeitungsanwendungen ist besonders die Lasermesstechnik an diesen Verteilungen interessiert. Sie werden in Messanordnungen, wie z.B. dem Laserlichtschnittverfahren, eingesetzt, aber auch in Bereichen des Alltags. Ein Beispiel dafür sind die Barcode-Scanner an Supermarktkassen.

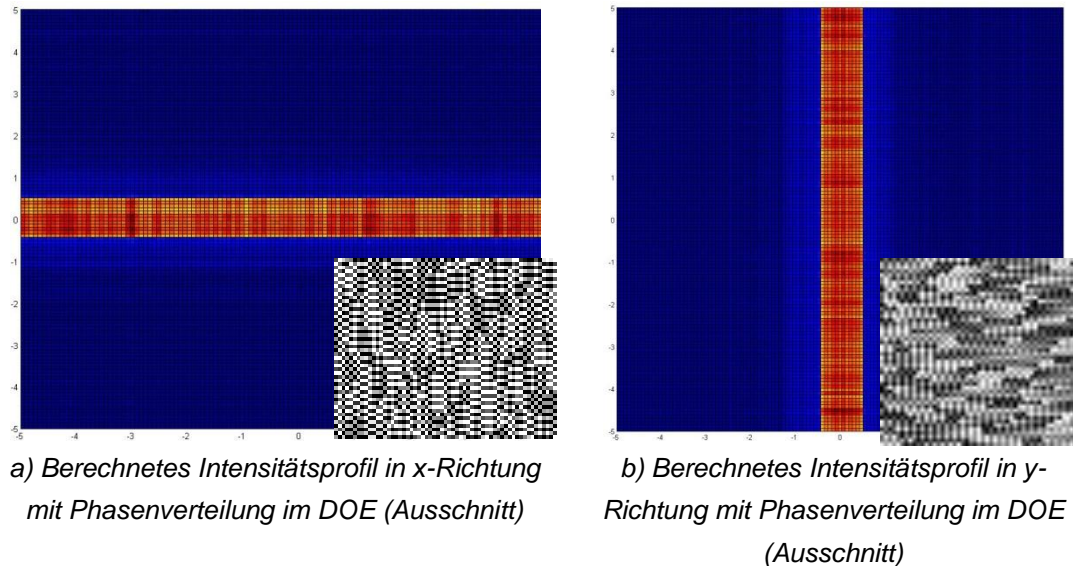


Abbildung 25 Simuliertes Ergebnis: Linienprofil in x- und y-Richtung (erstellt mit MATLAB)

Abbildung 25 zeigt zwei Varianten der Simulationsergebnisse. Die Linien sollen die gleiche Breite haben, mittig liegen und im betrachteten Bereich in ihrer Länge nicht begrenzt sein. Eine Linie sowohl in x- als auch in y-Richtung konnte erfolgreich separat erzeugt werden. Der hohe Kontrast und die exakte Breite des Profils bezeugen die guten Ergebnisse bei der DOE-Berechnung mit dem vorgestellten Algorithmus.

Die Ausschnitte aus den berechneten Phasenverteilungen der Elemente, die in Abbildung 25 zu sehen sind, geben einen Einblick in das Funktionsprinzip dieser beiden speziellen DOEs. Bei genauem Hinschauen sind kleine periodische, gitterartige Strukturen zu sehen, die jeweils in Ausrichtung der erzeugten Linie liegen. Die Beugungsmuster dieser „Elementargitter“ liegen jeweils senkrecht zur zu erzeugenden Linie. Dabei sind die Gitter so zueinander angeordnet, dass sich diese Beugungsmuster außerhalb des gewünschten Bereiches destruktiv, innerhalb des Bereiches jedoch konstruktiv überlagern. Tatsächlich ist eine ähnliche Herangehensweise als Multi Grid Array Stand der Technik (siehe 2.4.1).

4.1.3 Strahlformung: Hochschullogo

Nach den regelmäßigen Strukturen der letzten Abschnitte interessiert, ob auch Intensitätsprofile zu erzeugen sind, die diese Symmetrie nicht aufweisen. Wie schon in Abbildung 3 zu erkennen ist, sind Logos, die mittels DOE projiziert werden, gängige Anwendungen im Marketing- und Werbebereich. Somit soll das Logo der Hochschule Mittweida berechnet werden. Die gewünschte Verteilung wurde als Bitmap-Datei in den Algorithmus eingelesen und simuliert.

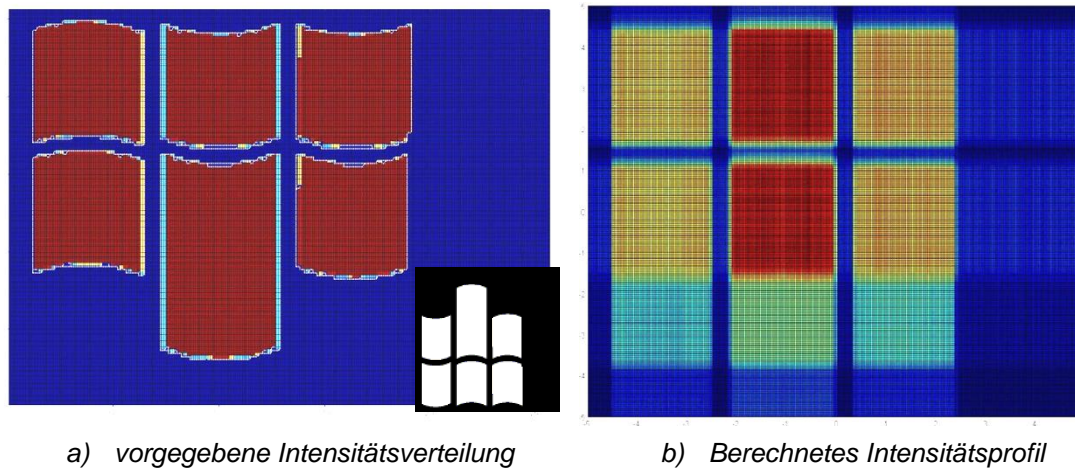


Abbildung 26 Simuliertes Ergebnis: Hochschullogo
(erstellt mit MATLAB)

Abbildung 26 offenbart, dass der Algorithmus nicht für jede beliebige Form der Zielintensität geeignet ist. Das Hochschullogo ist in der Simulation nicht als solches wieder zu erkennen. Die regelmäßige Anordnung der sechs Rechtecke kann zwar erzeugt werden, doch die Abweichungen in Form der Rundungen und des erhöhten oberen Fragments sind nicht möglich.

Auch eine Erhöhung der Rechengenauigkeit bringt keine nennenswerten Verbesserungen. Die dargestellte Intensitätsverteilung ist mit einer Diskretisierung von $N = 256$ Punkten je Raumrichtung und $p = 32$ Stufen der Phasenlage simuliert worden.

Weiter ist in Abbildung 26 zu sehen, dass das erzeugte Profil auf dem Kopf steht. Diese Spiegelung an der horizontalen Achse ist nicht bedingt durch den Algorithmus. Abbildung 26a zeigt das Zielprofil, wie es in MATLAB eingelesen wurde. Die Spiegelung erfolgt also schon während des Einleseprozesses. Die genauen programmtechnischen Hintergründe könnten für eine Weiterentwicklung des Programmmoduls erforscht werden. Es ist für die praktische Anwendung des Algorithmus jedoch nicht von Belang, ob das Bild auf dem Kopf steht, da es eine einfache Drehung des realen DOEs wieder korrekt ausrichten würde.

4.1.4 Strahlformung: Buchstabe M

Zur Überprüfung der Frage, ob der Algorithmus allgemein beliebig gestaltete Intensitätsverteilungen erzeugen kann und lediglich mit Rundungen Probleme hat, wird der Buchstabe „M“ in der Bildebene erzeugt.

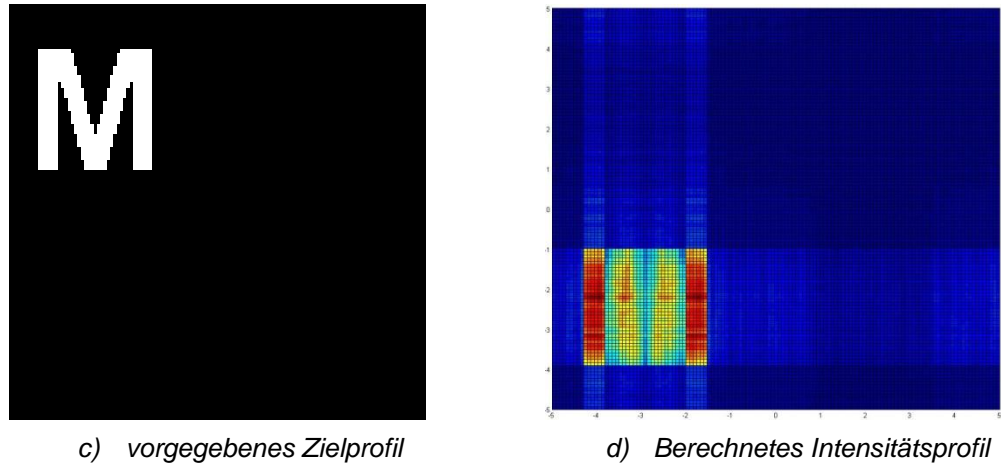


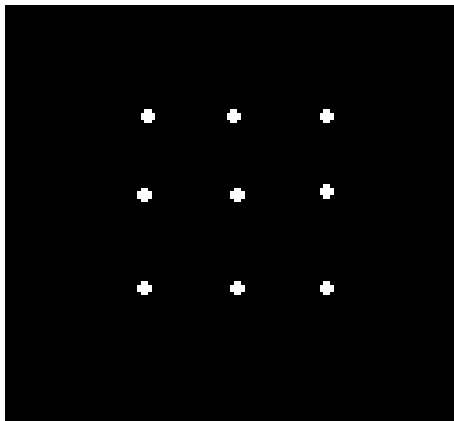
Abbildung 27 Simuliertes Ergebnis: Strahlformung "M"
(erstellt mit MATLAB)

Wie Abbildung 27 zeigt, gibt es auch Probleme bei der Darstellung allgemein beliebig geformter Strahlprofile. Der Buchstabe „M“, dessen Kontur komplett aus Geraden besteht, kann ebenfalls nicht simuliert werden. Das Ziel ist zwar zu erraten, doch nicht wirklich zu erkennen. Die bereits festgestellte Spiegelung an der x-Achse tritt auch hier auf. Um diese nochmals zu verdeutlichen, wurde das zu erzeugende „M“ in die obere linke Ecke gerückt. Eine physikalische Ursache für die Spiegelung würde auf beide Achsen wirken und das Bild in die untere rechte Ecke verschieben.

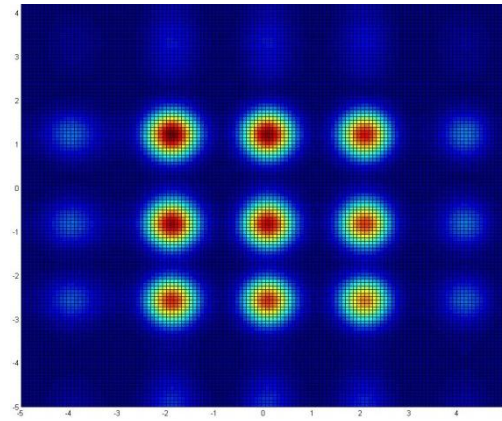
4.1.5 Strahlteiler: 3x3-Array

Die letzte untersuchte Anordnung ist ebenfalls im Bereich der Laserbearbeitung von Bedeutung. Zur Effizienzsteigerung werden Laserstrahlen oft in mehrere Teilstrahlen aufgeteilt, die z.B. die parallele Bearbeitung mehrerer Werkstücke ermöglichen. Meist wird dies über herkömmliche Strahlteilerspiegel realisiert. Eine neue Variante sind DOE-Strahlteiler.

Das Ziel besteht darin, den einfallenden Gaußstrahl in ein Raster von 3 x 3 erkennbar getrennter Einzelpeaks zu teilen. Diese lokalen Maxima sollten idealerweise alle die gleiche Intensitätsverteilung aufweisen. In Abbildung 28 sind eingelesenes Zielbild und das simulierte Ausgangsstrahlprofil dargestellt:



a) *Ideales Zielbild*



b) *Simuliertes Ergebnis*

Abbildung 28 Simuliertes Ergebnis: 3x3-Strahlteiler-Array
(erstellt mit MATLAB)

Die neun einzelnen Intensitätsmaxima sind deutlich zu sehen. Die Berechnung des geforderten Strahlteilers mit Hilfe des vorgestellten Algorithmus ist also grundsätzlich gut möglich.

In Abbildung 28b sind deutlich weitere Nebenordnungen der Beugung zu sehen. Diese kleinen Peaks lassen sich auch mit Erhöhung der Rechengenauigkeit nicht völlig unterdrücken. Da ihre Intensität jedoch deutlich niedriger ist, können sie für die Lasermaterialbearbeitung vernachlässigt werden. Die Schwellintensität z.B. für einen Materialabtrag wird außerhalb des gewünschten Arrays nicht überschritten.

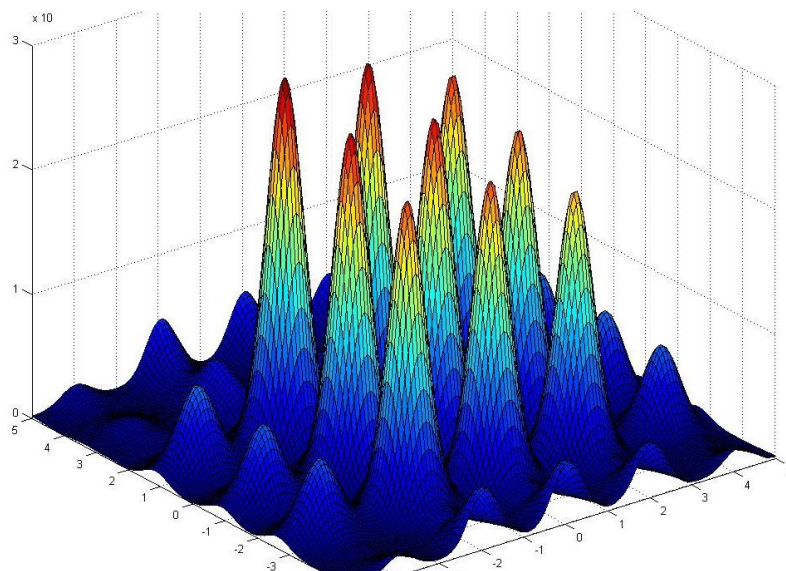


Abbildung 29 Strahlprofil des Fan-Out-Elements
(erstellt mit MATLAB)

Es ist weiter festzustellen, dass die Intensitäten der einzelnen Rasterpunkte nicht ideal gleich ist. In Abbildung 29 ist ein leichter Abfall von links oben nach rechts unten zu erkennen. Dieser ist nicht physikalisch typisch, sondern durch das numerische Verfahren bedingt. Der Startwert der zufälligen Phase beeinflusst hier das Ergebnis. Weitere Simulationen zeigen andere Verteilungen. Diese Intensitätsschwan-

kungen sollten für reproduzierbare Ergebnisse im praktischen Einsatz minimiert werden.

4.2 Überprüfung mit VirtualLab 4

Die berechneten Phasenverteilungen für die DOEs werden in LightTrans VirtualLab 4 eingelesen und Lichtdurchgang und Ausgangsverteilung simuliert. Die bereits erhaltenen Verteilungen werden einerseits bestätigt. Andererseits können sogar für DOEs, die zunächst als nicht gelungen eingestuft wurden, gute Ergebnisse nachgewiesen werden.

Die bisher simulierten Strahlverteilungen sind ausschließlich innerhalb des Algorithmus berechnet worden. Um die Ergebnisse objektiv beurteilen zu können, ist ihre Verifizierung durch weitere Instanzen notwendig. Der beste Beweis für die Funktionalität der berechneten DOEs wäre natürlich ihre praktische Herstellung und die Vermessung der realen Strahlverteilung.

Die Realisierung einer solchen Struktur aus 128x128 Pixeln, die alle über eine exakte Tiefe verfügen müssen, ist jedoch recht aufwändig. Aus diesem Grund soll zunächst eine theoretische Simulation des Strahlenganges durch ein solches DOE stattfinden. Die dafür genutzte Software ist das Optikdesign-Programm VirtualLab 4 der Firma LightTrans, Jena.

Die Simulation erfolgt mittels eines sogenannten Lightpath-Diagramms, in dem die wesentlichen Komponenten dargestellt sind. Abbildung 30 zeigt einen Screenshot des genutzten Lightpath-Modells, das im Wesentlichen aus einer gaußverteilten Strahlquelle („Gaussian Wave“), einer eingebetteten Funktion („Stored Function“) und einem virtuellen Schirm („Virtual Screen“) besteht. Die Funktion befindet sich 1 mm hinter der Strahlquelle, deren Wellen sich bis zur Transmissionsfunktion in Quarzglas ausbreiten. Dieser Modellaufbau ermöglicht die komplexe Simulation und die Berechnung, welche Verteilung sich auf dem Schirm einstellen würde.

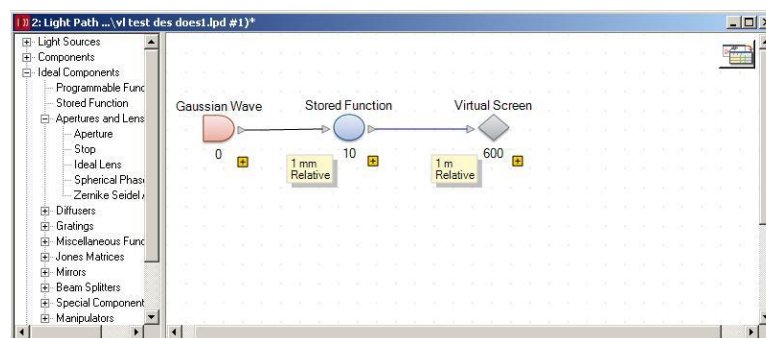


Abbildung 30 Screenshot: Lightpath-Diagramm des verwendeten Modells in VirtualLab 4

Genau genommen, simuliert dieses Modell nur den Strahlendurchgang durch irgendein DOE, das die einfallende Welle mit genau der vorher bestimmten Phasenänderung moduliert. Es wird kein konkretes DOE mit einem berechneten Höhenprofil gemäß Abbildung 2 simuliert, sondern nur dessen Wirkung. Die individuelle Phasenverteilung wurde über die Importfunktion von VirtualLab 4 aus den generierten Bitmap-Dateien und einige programminterne Transformationen gewonnen.

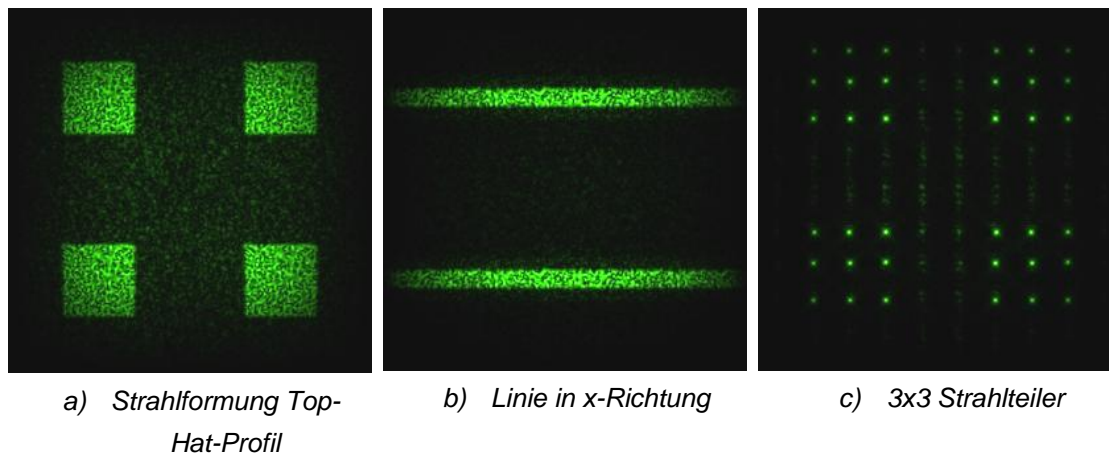


Abbildung 31 mit VirtualLab 4 simulierte Intensitätsverteilungen

In Abbildung 31a bis c sind die auf diese Art simulierten Intensitätsverteilungen von einigen der Teststrukturen dargestellt. Diese bestätigen die Richtigkeit der unter 4.1 erhaltenen Ergebnisse und zeigen die generelle Eignung des iterativen FOURIER-Transformations-Algorithmus, einen Gaußstrahl in die abgebildeten Verteilungen zu transformieren. Alle berechneten DOEs zeigen sehr gut erkennbar die gewünschten Bilder.

Im Gegensatz zu den simulierten Verteilungen mit MATLAB entstehen hier jedoch mehrere Bilder. Die Ursache dafür wurde aufgrund der Form einer zweidimensionalen optischen Faltung in einer Überlagerung des DOEs vermutet. Wie in den Vorüberlegungen ausgeführt (siehe 2.3.4), lässt eine Faltung in der FOURIER-Ebene (dem virtuellen Schirm) auf eine Multiplikation zweier Funktionen in der Beugungsebene (dem DOE) schließen. Die Form der vermeintlichen Faltung legt den Schluss nahe, dass die erste Funktion die Phasenverteilung des berechneten DOEs ist, die das gewünschte Beugungsbild generiert. Die zweite Funktion ist demzufolge eine zweidimensionale Deltafunktion mit verhältnismäßig kleinen Abständen. Diese Überlegungen führen zu dem Schluss, dass die zusätzlichen Ordnungen durch die diskrete Abtastung der ermittelten Funktion entstehen.

Im Laufe der Untersuchung wurde festgestellt, dass der Fehler in einer zusätzlichen Diskretisierung des DOEs liegt. Zur Erhöhung der Auflösung wurde das erhaltene Ergebnis in VirtualLab 4 so interpoliert, dass aus einem Pixel jeweils vier Pixel derselben Phasenlage wurden. Das heißt, jeder dieser Pixel wird viermal abgetastet, was als Multiplikation mit einer Deltafunktion beschrieben werden kann. Dies ist die

Ursache für exakt vier weitere Beugungsbilder. In einer weiteren Simulation ist dies beispielhaft korrigiert (siehe Abbildung 32).

Die genaue Klärung der Zusammenhänge und die Unterdrückung dieser zusätzlichen Ordnungen kann Inhalt weiterführender Untersuchungen sein. Ein aus den vorangegangenen Überlegungen resultierender Ansatz ist einfach die Verringerung der Pixelgröße des DOEs.

Die Qualität der erzeugten Verteilungen wird zudem durch das Auftreten von Specklemustern und deutlichem Rauschen gemindert. Die Rauschverluste, in Abbildung 31a besonders gut zu erkennen, führen zu einer Verringerung der Konversionseffizienz. Die bereits erwähnten Nebenmaxima, die bei der Erzeugung des 3x3-Strahlteilerarrays entstehen, sind auch in Bild c gut zu erkennen.

Die Simulation der unregelmäßigen Strahlverteilungen zeigt überraschenderweise sehr ansprechende Resultate. Im Gegensatz zu den prognostizierten Beugungsbildern, die die Ausgangsbilder schwer erkennen lassen, weisen die berechneten Intensitätsverteilungen die gewünschte Form aus (siehe Abbildung 32). Auch hier sind Speckle und Rauschanteile vorhanden. Die Mehrfachabbildung ist jedoch bereits korrigiert.

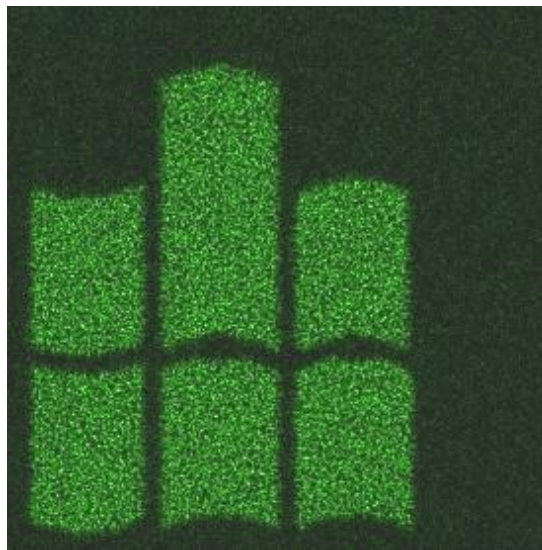


Abbildung 32 mit VirtualLab 4 simulierte Intensitätsverteilung des Hochschullogos

Die durchgeführten Simulationen lassen den Schluss zu, dass der iterative FOURIER-Transformations-Algorithmus auch bei unregelmäßigen Formen genutzt werden kann. Die errechneten Beugungsbilder der restlichen Strukturen sind im Anhang zu finden.

4.3 Auswertung

Die Ergebnisse zeigen, dass die Anforderungen für die Anwendungen Strahlformung und Strahlteilung durch die berechneten DOEs gut erfüllt werden. Der Algorithmus ist allgemein dafür geeignet. Dennoch müssen gewisse Einschränkungen gemacht werden.

4.3.1 Anwendung Strahlformung

Wie die vorangegangenen Abschnitte überzeugend darlegen, ist die Strahlformung eines gaußverteilten Laserstrahls in nahezu jede beliebige andere Intensitätsverteilung mittels des umgesetzten IFTA grundsätzlich gut realisierbar. Die Ergebnisse zeigen, dass Richtungsabhängigkeiten für die gewünschten Profile ausgeschlossen werden können. Prinzipiell wäre dies auch unlogisch, da durch Drehung des DOEs um 90° die exakt senkrechte Verteilung praktisch erzielt werden kann.

Es bereitet jedoch offenbar Schwierigkeiten, Intensitätsprofile zu realisieren, die Rundungen enthalten. Dies wird zunächst am Beispiel des Hochschullogos deutlich. Zum anderen zeigt Abbildung 33 einen weiteren Versuch, eine Halbkugelförmige Strahlverteilung zu generieren. Auch wenn eine Annäherung an das gewünschte Profil erreicht werden kann, sind doch deutliche Tendenzen zu einer rechteckigen Verteilung zu sehen.

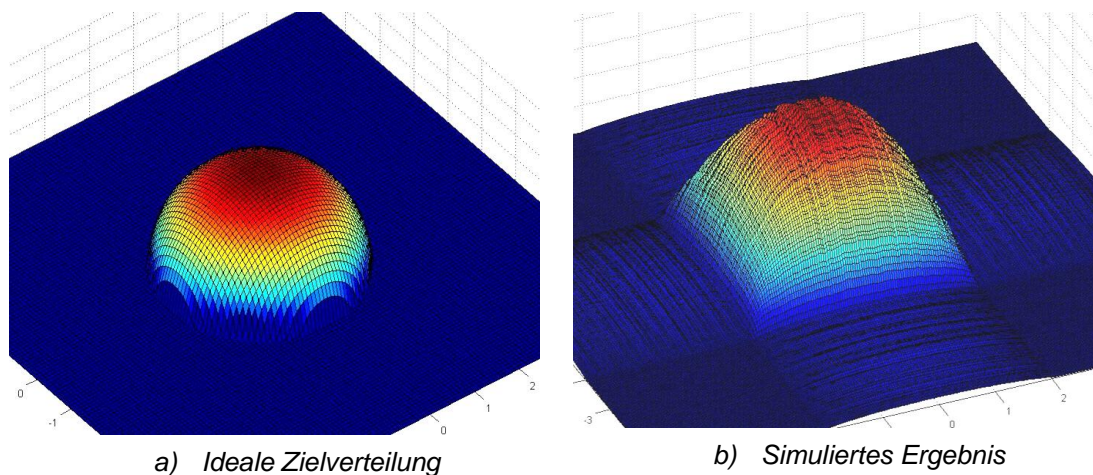


Abbildung 33 simuliertes Ergebnis: Halbkugelverteilung
(erstellt mit MATLAB)

Es muss allerdings eingeräumt werden, dass die Simulation der Ausgangsverteilung im Fall des Hochschullogos zu davon abweichenden Ergebnissen führt. Auch das zunächst negativ bewertete DOE zur Erzeugung des „M“, weist in der Nachberechnung durch VirtualLab 4 eine sehr viel bessere Verteilung auf (siehe Anhang 3). Die Resultate für Strahlformungsanwendungen sollten daher kritisch betrachtet werden. Es ist hier mit Nachdruck darauf hinzuweisen, welche große Bedeutung das Nachrechnen des erzeugten Strahlprofils für das Optikdesign von DOEs hat.

Alle errechneten DOEs weisen in ihrer Rekonstruktion signifikante Rauschanteile, deutliche Specklemuster und höhere Beugungsordnungen auf. Hier besteht für zukünftige Untersuchungen noch Optimierungspotential, z.B. durch nachgelagerte Algorithmen.

An dieser Stelle sei nochmals betont, dass der verwendete Algorithmus bei mehrmaligem Ausführen niemals zwei identische Ergebnisse erzielt. Durch die zufällig verteilte Startphasenlage kann ein mehrmaliges Durchlaufen des IFTA das Ergebnis auch optimieren. Dies gilt für alle durchgeführten Untersuchungen.

4.3.2 Einfluss des Eingangsstrahlradius

Unter dem Aspekt der Strahlformung wird in einer gesonderten Untersuchung das Augenmerk auf den Einfluss des Eingangsstrahls gerichtet. Eine umfangreiche Behandlung dieses Themas könnte in weiterführenden Experimenten erfolgen. An dieser Stelle wird nur der Einfluss des Parameters Eingangsstrahlradius des gaußverteilten Strahls betrachtet.

Der Einfluss des Eingangsstrahls muss auf zwei verschiedene Weisen betrachtet werden: vor und nach der DOE-Berechnung. Vor der Ermittlung des DOEs wird der entsprechende Radius in das Ergebnis des DOEs mit einfließen. Der Algorithmus wird versuchen, dieses bestmöglich auf die Eingangsverteilung anzupassen. Wie Abbildung 34 am Beispiel des „M“ zeigt, ist dies jedoch nur begrenzt möglich. Wird der Strahl zu schmal, ist eine ordentliche Rekonstruktion des Bildes nicht mehr möglich. Da die Pixelgröße des DOE nicht angepasst wird, stehen mit kleiner werdendem Strahl viel weniger Pixel zur Modulation zur Verfügung. Dies resultiert in den dargestellten Ausgangsintensitäten, bei denen sogar teilweise noch einzelne Beugungsordnungen des Strahls sichtbar sind.

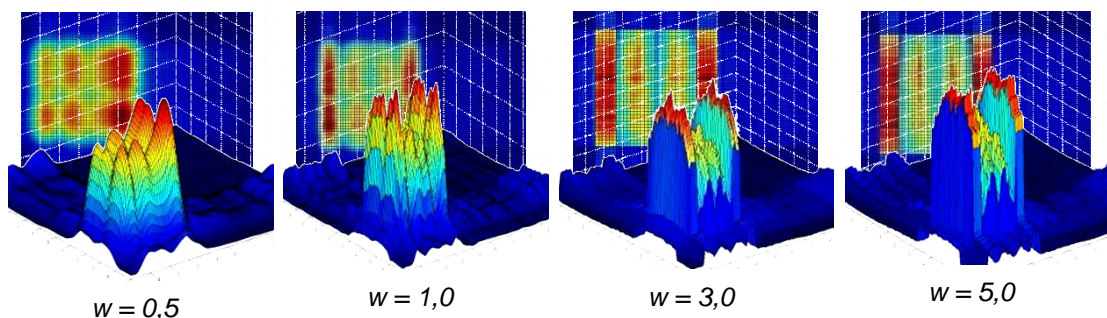


Abbildung 34 simuliertes Ergebnis: Strahlformung "M" bei verschiedenen Eingangsradien (erstellt mit MATLAB)

Mit zunehmendem Strahlradius können diese Effekte reduziert werden, bis sich keine Verbesserung mehr einstellt. Ein übermäßig breiter Strahl überdeckt das ganze DOE und wirkt fast schon, wie eine konstante Amplitudenverteilung.

Im vorliegenden Beispiel wird der angegebene Radius quantitativ zum berechneten Feld betrachtet. Es sind keine absoluten Größenangaben möglich und nötig. Bei einem Intervall von $[-5; 5]$ bedeutet ein Strahlradius für $1/e^2$ -Abfall von $w = 0,5$, dass nur ca. 1% der DOE-Fläche von rund 86% der Strahlleistung getroffen werden.

Die andere Art des Einflusses ist zu beachten, wenn das DOE bereits berechnet und hergestellt wurde. Das heißt, im Gegensatz zum eben betrachteten Fall ist die Form und Phasenmodulation des Elements fest und es wird nun erst mit unterschiedlich breiten Gaußstrahlen konfrontiert. Diese Untersuchung wurde mittels VirtualLab 4 durchgeführt. Die Ergebnisse sind in Abbildung 35 abgebildet.

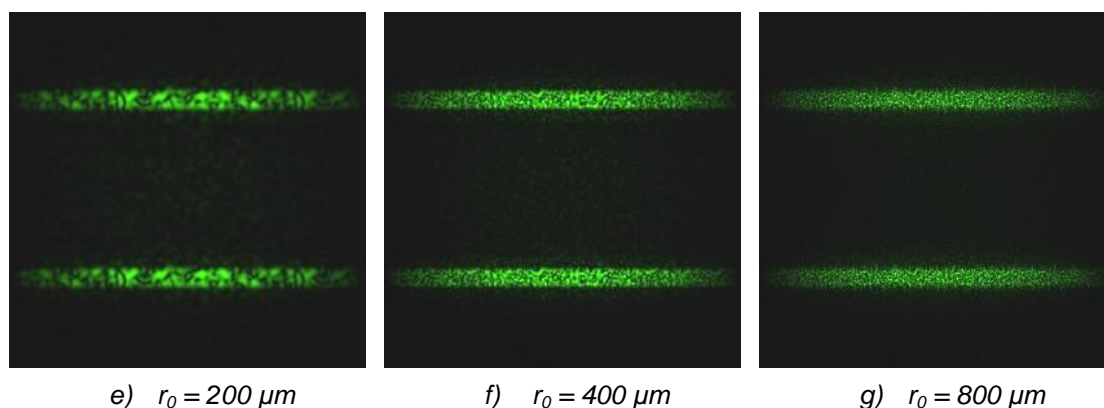


Abbildung 35 mit VirtualLab simuliert: Linienprofil bei verschiedenen Eingangsstrahlradien

Die Strahlradien sind auch hier bei $1/e^2$ -Abfall angegeben und sind diesmal als absolute Größen zu verstehen. Wie in 3.1.2 erläutert wurde, sind alle relativen Größen des Berechnungsprozesses mit Festlegung der Wellenlänge auf $\lambda = 532 \text{ nm}$ in VirtualLab 4 zu absolut geworden. Das DOE wird dann mit einer Pixelgröße von $13 \mu m$ festgelegt und hat damit die Gesamtkantenlänge von $1,664 \text{ mm}$.

Wie in Abbildung 35 zu sehen ist, hat die geänderte Strahlquelle sichtbaren Einfluss auf das rekonstruierte Bild. Die Feinheit der Speckle nimmt mit größerem Strahlradius ebenso zu, wie der Kontrast. Dies kann mit Hilfe der FOURIER-Optik erklärt werden.

Da diesem Prinzip zufolge das DOE die FOURIER-Transformierte des Beugungsbildes ist (bzw. die Inverse), kann die Betrachtung der Ortsfrequenzen hier angewandt werden. Voraussetzung ist natürlich, dass die Strahlen als exakt mittig angesehen werden, was hier gegeben ist. Der besonders schmale Gaußstrahl beleuchtet nur die inneren, niedrigen Ortsfrequenzen. Da diese für grobe Strukturen innerhalb eines Bildes verantwortlich sind, erklärt dies den geringen Kontrast und die sehr groben Specklemuster in Bild a. Desto größer der Strahlradius wird, umso mehr Frequenzen tragen zur Bildentstehung bei und sorgen für höhere Kontraste und feinere Specklemuster.

4.3.3 Anwendung Strahlteilung

In den vorderen Abschnitten wurden bereits die guten Ergebnisse beim Erzeugen von Strahlteilern unter Beweis gestellt. Sowohl MATLAB als auch VirtualLab 4 bestätigen die Eignung des IFTA für diese Aufgabenstellungen.

Wie die nachfolgende Abbildung 36 zeigt, lassen sich mit dieser Technik auch beliebige Verteilungen erzeugen. In diesem Fall wird ein 4x5-Array aus nahezu gleich hohen Peaks simuliert. Die mittels VirtualLab ermittelte Ausgangsverteilung ist in Anhang 3 einzusehen. Theoretisch sollte es auch keine Probleme bereiten, in einem solchen Array den Maxima verschiedene Werte zuzuordnen und so die Verteilung individueller zu steuern. Abbildung 36b zeugt vom Versuch, eine 3x3-Matrix durch „Ausschalten“ zweier Maxima unregelmäßig zu variieren.

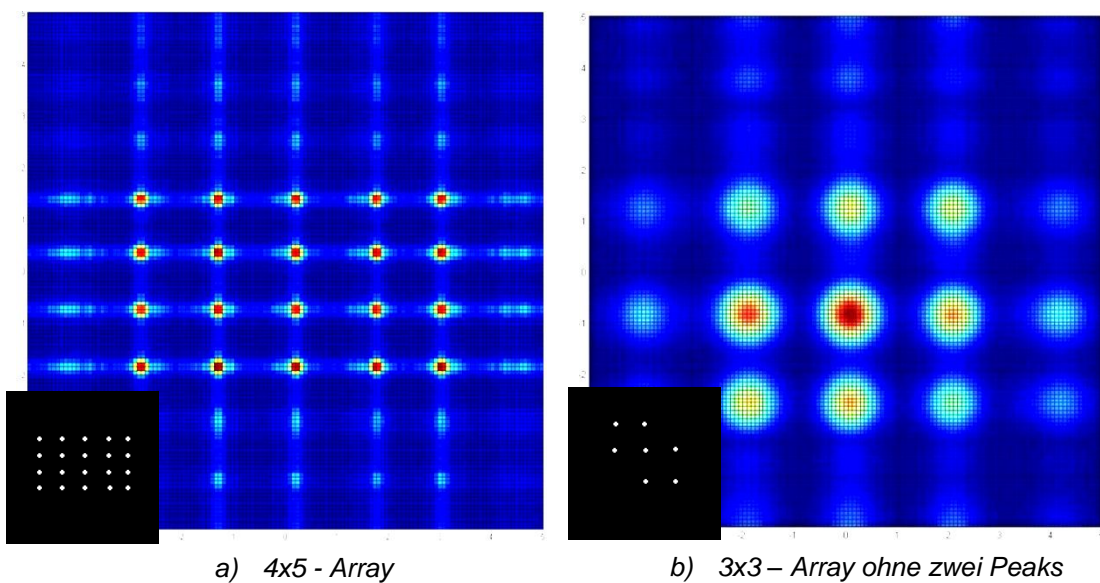


Abbildung 36 simulierte Ergebnisse: verschiedene Strahlteiler-Arrays
(erstellt in MATLAB)

Offenbar bereiten die niemals vermeidlichen Nebenmaxima bei einem Aufbrechen der absoluten Symmetrie dem IFTA erhebliche Probleme. Die gewünschte Verteilung ist nur ansatzweise zu erkennen. Allerdings zeigt sich auch, dass manche Endergebnisse von der Prognose in MATLAB abweichen können (siehe Hochschullogo). Diese Überprüfung konnte aus Zeitgründen leider nicht durchgeführt werden, so dass dies Gegenstand nachfolgender Versuche bleiben muss.

5 Zusammenfassung

Die vorgelegte Master Thesis dokumentiert die durchgeführten Untersuchungen zur Berechnung von DOEs. Sie stellt die erforderlichen physikalischen Grundlagen vor und präsentiert verschiedene existierende Algorithmen, die ihren Nutzen zur DOE-Berechnung bereits unter Beweis gestellt haben.

Der daraus ausgewählte iterative FOURIER-Transformations-Algorithmus (IFTA) wird erfolgreich in ein Programmmodul für MATLAB implementiert und mittels GUI für Berechnungen zugänglich gemacht. Es werden die einzelnen Schritte zur Entwicklung des Moduls, sowie zur Realisierung in MATLAB dargestellt und begründet.

Anschließend werden einige Beispiel-DOEs berechnet, die verschiedene in der Laserbearbeitung relevante Strahlprofile erzeugen können sollen. Es können erfolgreich fast alle geforderten Intensitätsverteilungen aus einem Gaußstrahl erzeugt werden. Die Ergebnisse werden anschließend im Optikdesign- und -analyseprogramm VirtualLab 4 verifiziert und die zu erwartenden generierten Bilder simuliert. Dies bestätigt die zuvor gewonnenen Erkenntnisse, dass der IFTA geeignet ist, DOE-Strukturen zu ermitteln.

Ausblickend können eine Unzahl von Erweiterungen realisiert werden. Neben den bereits im Kontext erwähnten Ansätzen sollte vor allem die Zuverlässigkeit und Flexibilität des Programmmoduls erweitert werden. Mögliche Ansätze sind weitere Algorithmen, die auch andere DOE-Aufgaben lösen können, die Implementierung anderer Strahlquellen oder das Hinzufügen weiterer Verfahren, die die Konversionseffizienz der erzeugten DOEs maßgeblich erhöhen.

Ein besonderes Augenmerk sollte auch auf der Diskretisierung der erzeugten Phasenverteilung in größere Pixel liegen. In einigen Versuchen konnte zwar die Diskretisierung prinzipiell realisiert werden, doch der Algorithmus konnte in diesen Fällen keine Lösungen mehr finden. Ein denkbarer Ansatz ist die sogenannte Soft Quantization, die von Frank Wyrowski vorgestellt wurde. Auch die aktuelle Einschränkung auf reine Phasen-DOEs sollte mit Erweiterungen überwunden werden können.

Der Vergleich mit dem internationalen Stand der Technik zeigt, dass all dies nur erste Schritte sein können, um den aktuellen Stand in der Berechnung diffraktiver optischer Elemente zunächst zu erreichen. Eine enge Zusammenarbeit von Physikern, Informatikern und Mathematikern könnte beste Ergebnis dabei erzielen.

Literatur

Literaturangaben in alphabetischer Reihenfolge des/der Autors/Autoren:

- [Bernert] Bernert, Cordula: Einführung in die Numerische Mathematik (Vorlesungsskript), Hochschule Mittweida, 2011
- [Brigham] Brigham, E. Oran: FFT – Schnelle Fourier-Transformation, München Wien, R. Oldenbourg Verlag, 1992
- [Brinkmann] Brinkmann, Matthias; Braun, Franz-Joseph; Hagemann, Malte; Jäger, Erwin: Diffraktive Optische Elemente erobern den Markt, Optik & Photonik, Nr.1 April 2009, Weinheim, Wiley-VCH Verlag GmbH & Co. KGaA, 2009
- [Ersoy] Ersoy, Okan K.: Diffraction, Fourier optics and Imaging, Hoboken (US-NJ), Verlag John Wiley & Sons, Inc., 2007
- [Fienup] Fienup, J.R.: Phase retrieval algorithms: a comparison, Applied Optics Vol 21, No. 15, S. 2758-2769, Washington, DC, Optical Society of America, 1982
- [Fricke] Fricke-Begemann, Thomas; Ihlemann, Jürgen: Direct light-coupling to thin-film waveguides using a grating-structured GRIN lens, Optics Express, Vol. 18, No. 19, S. 19860-19866, Washington, DC, Optical Society of America, 2010
- [Guessous] Guessous, Fouad: Photochemisch strukturierte computergenerierte Hologramme in Bakteriorhodopsin-Schichten (Dissertation), Marburg/Lahn, Phillips-Universität Marburg, Fachbereich Chemie, 2005

- [Huygens] Huygens, Christian; Lommel, E. (Hrsg.): Abhandlung über das Licht,
Leipzig, Verlag Wilhelm Engelmann, 1890
- [Kaakkunen] Kaakkunen, J.J.J.; Bekesi, J.; Ihlemann, J.; Simon, P.: Ablation of microstructures applying diffractive elements and UV femtosecond laser pulses,
Applied Physics A, No. 101, S. 225-229,
Berlin Heidelberg, Springer-Verlag, 2010
- [Kim] Kim, Hwi; Yang, Byungchoon; Lee, Byoungcho: Iterative Fourier transform algorithm with regularization for the optimal design of diffractive optical elements,
Journal Optical Society of America, Vol. 21, No. 12, S. 2353-2365,
Washington, DC., Optical Society of America, 2004
- [Lipson] Lipson, Stephen; Lipson, Henry; Tannhauser, David: Optik,
Berlin Heidelberg, Springer-Verlag, 1997
- [Löchel] Löchel, Bernd; Schondelmaier, Daniel: Information Sheet - Diffractive Optical Elements,
Berlin, Anwenderzentrum für Mikrotechnik, 2010
verfügbar unter: <http://www.azm.bessy.de>
- [O'Shea] O'Shea, Donald C.; Suleski, Thomas J.; Kathman, Alan D.; Prather, Dennis W.: Diffractive Optics – Design, Fabrication, and Test,
Bellingham (US-WA), SPIE – The Society of Photo-Optical Instrumentation Engineers, 2004
- [Pawlowski] Pawlowski, E.; Dobschal, H.-J.; Falkenstörfer, O.; Fürst, W.; Steiner, R.; Steingrüber, R.: Multi-wavelength optical pickup with diffractive optical elements,
Journal of Modern Optics, Vol. 47, No. 13, S. 2437 – 2453,
Verlag Taylor & Francis Ltd., 2000

- [Pedrotti] Pedrotti, Frank L.; Pedrotti, Leno S.; Bausch, Werner; Schmidt, Hartmut: Optik für Ingenieure, Berlin Heidelberg, Springer-Verlag, 2002
- [Ripoll] Ripoll, Olivier; Kettunen, Ville; Herzig, Hans-Peter: Review of iterative Fourier-transform algorithms for beam shaping applications, Optical Engineering 43, issue 11, S. 2549-2556, 2004
- [Schwetlick] Schwetlick, Hubert; Kretzschmar, Horst: Numerische Verfahren für Naturwissenschaftler und Ingenieure, Leipzig, Fachbuchverlag Leipzig, 1991
- [Sommerfeld] Sommerfeld, Arnold: Mathematische Theorie der Diffraction, Göttingen, Mathematische Annalen XLVII, 1896
- [Stankovic] Stankovic, Svetomir: Diffraktive Optische Elemente für moderne mikrooptische Systeme (Dissertation), Darmstadt, Technische Universität Darmstadt, Fachbereich Physik, 2002
- [Stoer] Stoer, Josef: Numerische Mathematik 1 (6. Auflage), Berlin Heidelberg, Springer-Verlag, 1993
- [Träger] Träger, Frank (Ed.): Springer Handbook of Lasers and Optics, Berlin Heidelberg, Springer-Verlag, 2007
- [Völkel] Völkel, Reinhard; Zimmermann, Maik: Homogenisierung von Laserstrahlen, Photonik 3/2006, S. 76 – 79, Fellbach, AT-Fachverlag GmbH, 2006
- [Wyrow2006] Wyrowski, Frank; Schimmel, Hagen: Elektromagnetisches Optikrechnen – eine Einführung, Photonik 6/2006, S. 54-57, Fellbach, AT-Fachverlag GmbH, 2006

[Wyrow2007] Wyrowski, Frank; Schimmel, Hagen; Kuhn, Michael: Elektromagnetisches Optikrechnen – Modellierung realer Lichtquellen, Photonik 2/2007, S. 66-70, Fellbach, AT-Fachverlag GmbH, 2007

Anlagen

Anlage 1 – Struktogramm des FFT-Algorithmus.....	A-1
Anlage 2 – Code des MATLAB-Programmmoduls.....	A-3
Anlage 3 – Bilder.....	A-8

Anlage 1 – Struktogramm des FFT-Algorithmus

erarbeitet aus dem mittels Flussdiagramm dargestellten Algorithmus in [Brigham], S. 195

Eingabewerte: N, \vec{f}

Initialisierungswerte: $N2 = N/2$; $\gamma = \log_2(N)$				
Schleife: für $l = 1$ bis γ (läuft Knotenspalten ab)				
	Schleife: für $I = 0$ bis $(2^{l-1}-1)$ (läuft duale Knotenpakete ab)			
		Schleife: für $n = (0+I*2*N2)$ bis $((N2-1)+I*2*N2)$		
			<u>Funktion: $p = p_Bestimm(n, \gamma, l)$</u>	
			$T1 = \exp(-i*2\pi*p/N)*f(n+1+N2)$ (Zwischenergebnis)	
			$f(n+1+N2) = f(n+1) - T1$	
			$f(n+1) = f(n+1) + T1$	
	$N2 = N2 / 2$			
	$f = f$ neu umgeordnet durch Bitumkehr der Indizes			
Grafische Ausgabe: f				

Unterfunktion $p_Bestimm$ (Übergabewerte: n, γ, l)

$n = n$ als Binärzahl um $(\gamma-l)$ nach rechts verschoben	
$p = 0$	
Schleife: für $i = \gamma-1$ bis 0 (läuft rückwärts)	
	$a = \text{Rest von } n/2 \text{ (muss entweder 0 oder 1 sein)}$
	$n = n$ als Binärzahl um 1 nach rechts verschoben
	$p = p + a * 2^i$
Ausgabewert: p	

$p_Bestimm$ berechnet den Exponenten in W^p durch verschiedene Bitoperationen an n .

Anlage 2 – Code des MATLAB-Programmmoduls

Funktion „Set_Matrix“

```
%Funktion Set_Matrix
%*****
%erstellt Matrix als Berechnungsgrundlage

%liest Parameter aus GUI aus

    %numerische Auflösung
    N = str2double(get(handles.edit4,'String'));

    %Gaußradius
    w = str2double(get(handles.edit1,'String'));

    %Spitzenintensität
    I0 = str2double(get(handles.edit2,'String'));

%Erstellen des x-Rasters und der Eingangs- und Ausgangsverteilung

x=(-5:10/(N-1):5);      %x-Achse als Zeilenvektor
y=(-5:10/(N-1):5)';     %y-Achse als Spaltenvektor
I = zeros(N,N,8);
for n=1:N
    I(n,:,1)=x;
    I(:,n,2)=y;
end
%Array enthält 2D-Arrays:
%1. Matrix      x-Werte
%2. Matrix      y-Werte
%3. Matrix      Zielfunktion Iout
%4. Matrix      Eingangsfunktion Iin
%5. Matrix      komplexe Amplitude A
%6. Matrix      komplexe Amplitude a
%7. Matrix      komplexe Amplitude a2
%8. Matrix      komplexe Amplitude A2
%Zu Beginn sind bis auf x- und y-Werte alle Werte 0

%liest ggf. Bild aus Datei ein (nur wenn entspr. Auswahl)
if get(handles.Wahl_bmp,'Value')==1,
    Bild = im2double(imread(get(handles.edit9,'String'),'bmp'));
%Bild wird eingeladen (Dimensionen prüfen!)
end

for n=1:N
    for m=1:N
        %Eingangsverteilung
        I(n,m,4)=I0*exp(-2*(I(n,m,1).^2+I(n,m,2).^2)./(w.^2));
    %Gauß-Verteilung

        %Ausgangsverteilung erstellen (abhängig von Auswahl)
```

```

    %Ausgangsverteilung Top-Hat
    if get(handles.Wahl_tophead, 'Value')==1,
        deltax=str2double(get(handles.edit6, 'String'));
        deltax=str2double(get(handles.edit7, 'String'));
        if -deltax/2 <= I(n,m,1) && I(n,m,1) <= deltax/2 &&
I(n,m,2) >= -deltay/2 && I(n,m,2) <= deltax/2
            I(n,m,3)=1;
        else
            I(n,m,3)=0;
        end
    end

    %Ausgangsverteilung Linien Top-Hat
    if get(handles.Wahl_linie, 'Value')==1,
        staerke=str2double(get(handles.edit8, 'String'));
        winkel=str2double(get(handles.edit13, 'String'));
        if winkel == 90, %Sonderfall 90°
            if I(n,m,1) >= -staerke/2 && I(n,m,1) <= staerke/2
                I(n,m,3)=1;
            else
                I(n,m,3)=0;
            end
        else %allgemein für Linie mit Stärke und
Winkel
            if I(n,m,2) >= tan(winkel*pi/180).*I(n,m,1)-
staerke/(2*cos(winkel*pi/180)) &&...
                I(n,m,2) <=
tan(winkel*pi/180).*I(n,m,1)+staerke/(2*cos(winkel*pi/180))
                I(n,m,3)=1;
            else
                I(n,m,3)=0;
            end
        end
    end

    %Ausgangsverteilung aus bmp eingeladen
    if get(handles.Wahl_bmp, 'Value')==1,
        I(n,m,3)=Bild(n,m)./255;
    end

    %komplexe Start-Amplitude (zufällige Phase)
    I(n,m,5)=sqrt(I(n,m,3))*exp(i*2*pi*rand);
end
end

save 'I.mat' I; %erstellt neue Startmatrix I
surf(handles.axes2,I(:, :, 1),I(:, :, 2),I(:, :, 3)); %Zeichnet Zielfeld in
axes2

```

Funktion „IFTA“

```

%Funktion IFTA
%*****
%führt iterativen Fourier-Transformations-Algorithmus aus

%Einlesen der Parameter
%numerische Auflösung

```

```

N = str2double(get(handles.edit4, 'String'));

%Phasendiskretisierung
p = str2double(get(handles.edit5, 'String'));

%Iterationen s
if get(handles.Wahl_Schritt, 'Value')==1,
    s=1;
else
    s = str2double(get(handles.edit3, 'String'));
end

load 'I.mat' I;      %Lädt Datenblock I aus Datei

for q=1:s,

%ifft durchführen
I(:, :, 6)=ifft2(I(:, :, 5));

%Extrahieren der Phase des Ergebnisses
I(:, :, 7)=angle(I(:, :, 6));

%Quantisieren der Phase in p Phasenstufen
I(:, :, 7)=I(:, :, 7)/(2.*pi).*p; %Normierung, so dass die Stufen ganzzah-
lig werden
I(:, :, 7)=round(I(:, :, 7));      %Rundet Phasenwerte auf diskrete p Stufen
I(:, :, 7)=I(:, :, 7).(2.*pi)./p; %Normierung, so dass die Stufen ganzzah-
lig werden

%Korrektur der Amplitude
I(:, :, 7)=sqrt(I(:, :, 4))./(N.^2).*exp(i.*I(:, :, 7));

%fft durchführen
I(:, :, 8)=fft2(I(:, :, 7));

surf(handles.axes1, I(:, :, 1), I(:, :, 2), (abs(I(:, :, 8)))^2);

I(:, :, 5)=sqrt(I(:, :, 3)).*exp(i.*angle(I(:, :, 8)));

end

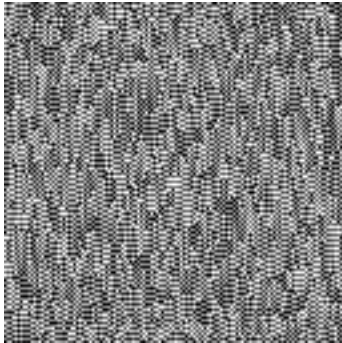
save 'I.mat' I;      %Speichert Datenblock I in Datei

%Speichern der Phase in bmp-Datei
imwrite((angle(I(:, :, 7))+pi)/(2*pi), 'tmp.bmp', 'bmp');

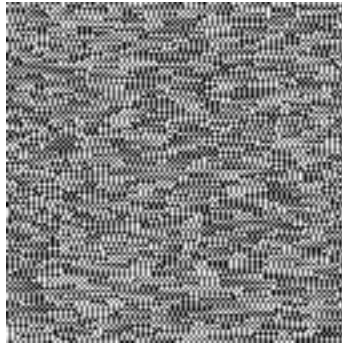
```


Anlage 3 – Bilder

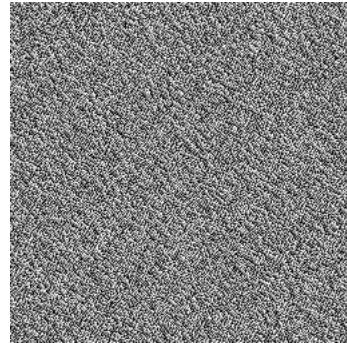
Bild A3 – 1a-f: Berechnete Phasenverteilungen für DOEs:



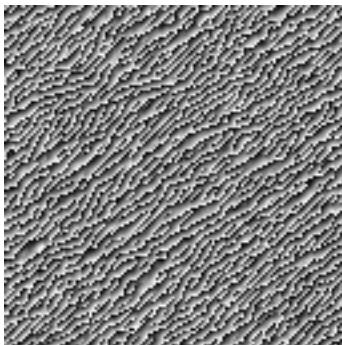
a) Linie in x-Richtung



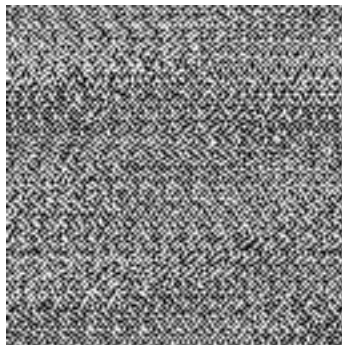
b) Linie in y-Richtung



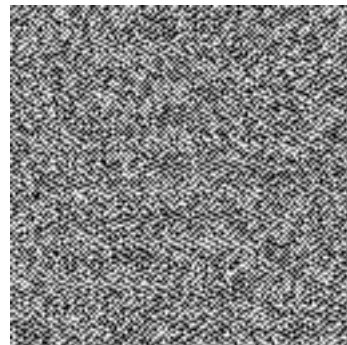
c) Hochschullogo



d) Strahlformer: M

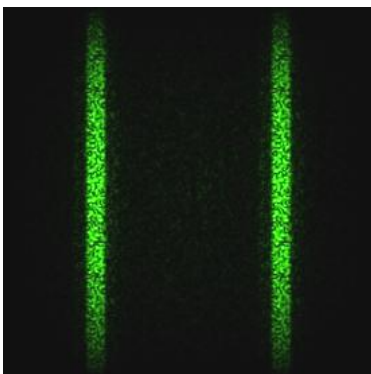


e) 3x3 Array

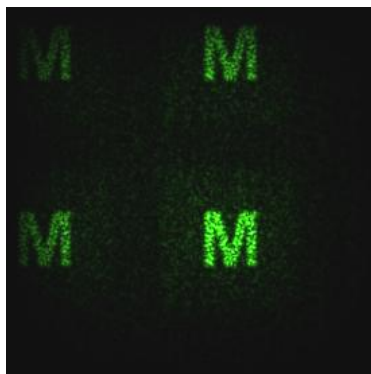


f) 4x5 Array

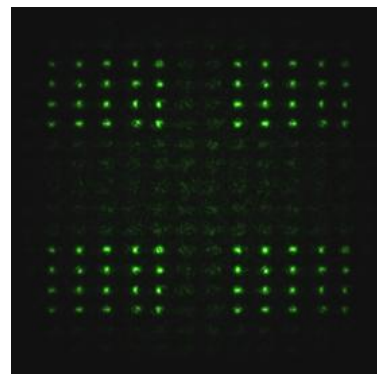
Bild A3 – 2a-c: Ausgangsbilder der DOEs, simuliert mit LightTrans VirtualLab 4:



h) Linie in y-Richtung



i) Strahlformung „M“



j) 4x5 Strahlteiler

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, den 15.10.2011

Falko Jahn